Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 194420)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (18-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Dec 18, 2003 this sequence version replaced gi:38259237.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L23552
Center clone name: 218_I_23
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 115229: contig of 115229 bp in length
* 115230 115329: gap of 100 bp
* 115330 194420: contig of 79091 bp in length.

FEATURES Location/Qualifiers
source 1. .194420
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-218L23"
/clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Alignment Scores:

Pred. No.:	374	Length:	194420
Score:	58.00	Matches:	11
Percent Similarity:	70.59%	Conservative:	1
Best Local Similarity:	64.71%	Mismatches:	5
Query Match:	58.59%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AC124133 (1-194420)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||       |||   |||||||||   |||||||||||||||   :::|||
Db      29103 ATGAAACTTCACAAAGTCCTTTCTATCCTTCTGTGGACACCCATTAGTCTC 29053
```

RESULT 8
AC096099/c
LOCUS AC096099 244669 bp DNA linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-27M23, WORKING DRAFT SEQUENCE.
ACCESSION AC096099
VERSION AC096099.6 GI:24941648
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 244669)

AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

```
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 244669)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 244669)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22772123.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GEKW
                Center clone name: CH230-27M23
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 215199 bases at least Q40
                Consensus quality: 218145 bases at least Q30
                Consensus quality: 219884 bases at least Q20
                Estimated insert size: 219668; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   244669: contig of 244669 bp in length.
FEATURES                Location/Qualifiers
     source             1. .244669
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-27M23"
     misc_feature       1. .1344
                        /note="wgs_contig"
     misc_feature       4991. .6114
                        /note="wgs_contig"
     misc_feature       92124. .93176
                        /note="wgs_contig"
     misc_feature       243409. .244669
                        /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:                485             Length:          244669
Score:                    58.00           Matches:         9
Percent Similarity:       81.25%          Conservative:    4
Best Local Similarity:    56.25%          Mismatches:      3
Query Match:              58.59%          Indels:          0
DB:                       2               Gaps:            0

US-10-799-747-116 (1-20) x AC096099 (1-244669)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAla 16
              |||   |||||||||||   ||||||::::::|||   |||::::::
Db     112297 ATGTGTGCACACTCTGTGTTTTCTTTTGTGATTTGGTTACCTCACTCA 112250

RESULT 9
AC140313/c
LOCUS       AC140313               245830 bp    DNA     linear   HTG 23-FEB-2003
DEFINITION  Mus musculus chromosome UNK clone RP24-324C6, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC140313
VERSION     AC140313.1  GI:28475563
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 245830)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245830)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information --------------
            Center project name: M_BB0324C06
            ------------------------------------------------
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 242982 bases at least Q40
            Consensus quality: 243412 bases at least Q30
            Consensus quality: 243721 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
            * be preserved.
            *        1     1085: contig of 1085 bp in length
            *     1086     1185: gap of unknown length
            *     1186    11944: contig of 10759 bp in length
            *    11945    12044: gap of unknown length
            *    12045    30225: contig of 18181 bp in length
            *    30226    30325: gap of unknown length
            *    30326    63384: contig of 33059 bp in length
            *    63385    63484: gap of unknown length
            *    63485   117854: contig of 54370 bp in length
            *   117855   117954: gap of unknown length
            *   117955   166434: contig of 48480 bp in length
            *   166435   166534: gap of unknown length
            *   166535   245830: contig of 79296 bp in length.
FEATURES             Location/Qualifiers
     source          1. .245830
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-324C6"
     misc_feature    1. .1085
                     /note="assembly_name:Contig5"
     misc_feature    1186. .11944
                     /note="assembly_name:Contig8"
     misc_feature    12045. .30225
                     /note="assembly_name:Contig9"
     misc_feature    30326. .63384
                     /note="assembly_name:Contig10"
     misc_feature    63485. .117854
                     /note="assembly_name:Contig11"
     misc_feature    117955. .166434
                     /note="assembly_name:Contig12"
     misc_feature    166535. .245830
                     /note="assembly_name:Contig13"
ORIGIN

Alignment Scores:
Pred. No.:              488             Length:         245830
Score:                  58.00           Matches:        11
Percent Similarity:     70.59%          Conservative:   1
Best Local Similarity:  64.71%          Mismatches:     5
Query Match:            58.59%          Indels:         0
DB:                     2               Gaps:           0

US-10-799-747-116 (1-20) x AC140313 (1-245830)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||      |||   |||||||||   |||||||||||||||   :::|||
Db      14529 ATGAAACTTCACAAAGTCCTTTCTATCCTTCTGTGGACACCCATTAGTCTC 14479

RESULT 10
AC013568/c
LOCUS       AC013568              156471 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-1B9, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   AC013568
VERSION     AC013568.3  GI:7107832
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156471)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-1B9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156471)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6514007.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2468
                Center clone name: 1_B_9
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136271 bases at least Q40
                Consensus quality: 149734 bases at least Q30
                Consensus quality: 153894 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 155571; sum-of-contigs
                Quality coverage: 5.4 in Q20 bases; agarose-fp
                Quality coverage: 5.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1916: contig of 1916 bp in length
            *     1917     2016: gap of 100 bp
            *     2017     4993: contig of 2977 bp in length
            *     4994     5093: gap of 100 bp
            *     5094    10286: contig of 5193 bp in length
            *    10287    10386: gap of 100 bp
            *    10387    15563: contig of 5177 bp in length
            *    15564    15663: gap of 100 bp
            *    15664    27659: contig of 11996 bp in length
            *    27660    27759: gap of 100 bp
            *    27760    41302: contig of 13543 bp in length
            *    41303    41402: gap of 100 bp
            *    41403    54013: contig of 12611 bp in length
            *    54014    54113: gap of 100 bp
            *    54114    75521: contig of 21408 bp in length
            *    75522    75621: gap of 100 bp
            *    75622   109129: contig of 33508 bp in length
            *   109130   109229: gap of 100 bp
            *   109230   156471: contig of 47242 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156471
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-1B9"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1916
                     /note="assembly_fragment"
     misc_feature    2017. .4993
```

```
                     /note="assembly_fragment"
     misc_feature    5094. .10286
                     /note="assembly_fragment"
     misc_feature    10387. .15563
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    15664. .27659
                     /note="assembly_fragment"
     misc_feature    27760. .41302
                     /note="assembly_fragment"
     misc_feature    41403. .54013
                     /note="assembly_fragment"
     misc_feature    54114. .75521
                     /note="assembly_fragment"
     misc_feature    75622. .109129
                     /note="assembly_fragment"
     misc_feature    109230. .156471
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
ORIGIN

Alignment Scores:
Pred. No.:                444             Length:         156471
Score:                    57.00           Matches:        10
Percent Similarity:       81.25%          Conservative:   3
Best Local Similarity:    62.50%          Mismatches:     3
Query Match:              57.58%          Indels:         0
DB:                       2               Gaps:           0

US-10-799-747-116 (1-20) x AC013568 (1-156471)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              :::::::||||||   |||   ||||||   |||||||||||||:::|||
Db     146405 TCTAGTCATTCCCATCTTTGCTTTCTGAGCTGGACTCCTTACAGTCTG 146358

RESULT 11
AL358815/c
LOCUS       AL358815               192867 bp    DNA     linear   PRI 02-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-526J3 on chromosome 9, complete
            sequence.
ACCESSION   AL358815
VERSION     AL358815.12  GI:12584727
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192867)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 28, 2001 this sequence version replaced gi:12581015.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-526J3 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-526J3 The true
            left end of clone RP11-18A3 is at 135297 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .192867
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-526J3"
                     /clone_lib="RPCI-11.2"
     repeat_region   328. .681
                     /note="177 copies 2 mer aa 54% conserved"
     repeat_region   2745. .2993
                     /note="L2 repeat: matches 2143. .2387 of consensus"
     repeat_region   3071. .3179
                     /note="L1MD2 repeat: matches 5962. .6069 of consensus"
     repeat_region   3188. .3408
                     /note="L2 repeat: matches 1908. .2140 of consensus"
     repeat_region   3515. .3774
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   3775. .3844
                     /note="AluSx repeat: matches 220. .290 of consensus"
     repeat_region   4518. .4577
                     /note="L2 repeat: matches 1434. .1492 of consensus"
     repeat_region   4554. .4736
                     /note="L2 repeat: matches 2242. .2419 of consensus"
     repeat_region   4756. .4956
                     /note="MIR repeat: matches 2. .211 of consensus"
     repeat_region   4957. .5256
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   5257. .5309
                     /note="MIR repeat: matches 211. .261 of consensus"
     repeat_region   7073. .7408
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   7657. .7973
                     /note="AluJo repeat: matches 1. .286 of consensus"
     repeat_region   7979. .8189
                     /note="MIR repeat: matches 47. .252 of consensus"
     repeat_region   8329. .8463
                     /note="AluJo/FRAM repeat: matches 178. .311 of consensus"
     repeat_region   10397. .10739
                     /note="L1MB5 repeat: matches 4107. .4444 of consensus"
     repeat_region   10740. .11127
                     /note="L1MA2 repeat: matches 5685. .6079 of consensus"
     repeat_region   11128. .11429
                     /note="AluSx repeat: matches 5. .305 of consensus"
     repeat_region   11430. .11661
                     /note="L1MA2 repeat: matches 6079. .6308 of consensus"
     repeat_region   11662. .13015
                     /note="L1MB5 repeat: matches 4444. .5403 of consensus"
     repeat_region   12986. .13083
                     /note="L1MB5 repeat: matches 5994. .6096 of consensus"
     repeat_region   13099. .13391
                     /note="AluSc repeat: matches 1. .290 of consensus"
     repeat_region   13417. .13561
                     /note="AluJo/FRAM repeat: matches 154. .300 of consensus"
     repeat_region   14426. .14737
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   15741. .15784
                     /note="22 copies 2 mer aa 75% conserved"
     repeat_region   17546. .17778
                     /note="MIR repeat: matches 16. .262 of consensus"
     repeat_region   19580. .19879
                     /note="AluSq repeat: matches 1. .301 of consensus"
     repeat_region   20116. .20396
                     /note="AluSg repeat: matches 1. .281 of consensus"
```

```
repeat_region   21132. .21417
                /note="AluSq repeat: matches 5. .304 of consensus"
repeat_region   21828. .21879
                /note="MIR repeat: matches 47. .97 of consensus"
repeat_region   22385. .22689
                /note="AluSq repeat: matches 1. .308 of consensus"
repeat_region   22843. .23149
                /note="AluSx repeat: matches 1. .308 of consensus"
repeat_region   23335. .23431
                /note="MIR repeat: matches 158. .260 of consensus"
repeat_region   23380. .23441
                /note="L2 repeat: matches 2648. .2709 of consensus"
repeat_region   23697. .23837
                /note="FLAM_C repeat: matches 1. .133 of consensus"
repeat_region   23850. .24147
                /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region   24429. .24737
                /note="AluSq repeat: matches 1. .301 of consensus"
repeat_region   25297. .25354
                /note="MIR repeat: matches 56. .120 of consensus"
misc_feature    26523. .27318
                /note="CpG island"
                /evidence=not_experimental
repeat_region   27696. .27925
                /note="MIR repeat: matches 6. .241 of consensus"
repeat_region   27923. .28001
                /note="MER5B repeat: matches 6. .87 of consensus"
repeat_region   28069. .28186
                /note="MIR repeat: matches 67. .182 of consensus"
repeat_region   28585. .29048
                /note="LTR36 repeat: matches 107. .612 of consensus"
repeat_region   29049. .29337
                /note="AluSq repeat: matches 1. .296 of consensus"
repeat_region   29338. .29451
                /note="LTR36 repeat: matches 1. .107 of consensus"
repeat_region   29947. .30256
                /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region   30780. .30953
                /note="L2 repeat: matches 2513. .2710 of consensus"
repeat_region   31789. .31966
                /note="MIR repeat: matches 34. .221 of consensus"
repeat_region   31967. .32242
                /note="AluSx repeat: matches 1. .276 of consensus"
repeat_region   32243. .32257
                /note="MIR repeat: matches 221. .234 of consensus"
repeat_region   32405. .32548
                /note="L2 repeat: matches 2374. .2509 of consensus"
repeat_region   32765. .32967
                /note="MIR repeat: matches 20. .235 of consensus"
repeat_region   33210. .33501
                /note="AluSx repeat: matches 21. .309 of consensus"
repeat_region   34265. .34555
                /note="AluSx repeat: matches 1. .292 of consensus"
repeat_region   34621. .34759
                /note="AluJo repeat: matches 1. .139 of consensus"
repeat_region   34877. .35100
                /note="L2 repeat: matches 2065. .2332 of consensus"
repeat_region   35112. .35449
                /note="MLT1D repeat: matches 12. .376 of consensus"
repeat_region   35450. .35759
                /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region   35760. .35891
                /note="MLT1D repeat: matches 376. .501 of consensus"
repeat_region   36022. .36391
                /note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region   36553. .36670
                /note="L2 repeat: matches 2582. .2699 of consensus"
repeat_region   36672. .36837
                /note="FRAM repeat: matches 2. .176 of consensus"
repeat_region   38859. .38936
                /note="MER47A repeat: matches 293. .366 of consensus"
repeat_region   38976. .39054
                /note="MER47 repeat: matches 1. .79 of consensus"
repeat_region   39284. .39594
                /note="AluY repeat: matches 1. .309 of consensus"
repeat_region   39959. .40181
                /note="MER58 repeat: matches 1. .229 of consensus"
repeat_region   40176. .40253
                /note="MER58 repeat: matches 2214. .2420 of consensus"
repeat_region   40954. .41068
                /note="MIR repeat: matches 117. .231 of consensus"
repeat_region   41524. .41559
                /note="L2 repeat: matches 2711. .2746 of consensus"
repeat_region   41699. .41792
                /note="L2 repeat: matches 2407. .2505 of consensus"
repeat_region   42767. .43070
                /note="AluSx repeat: matches 1. .303 of consensus"
repeat_region   43268. .43568
                /note="AluSq repeat: matches 1. .298 of consensus"
repeat_region   43854. .44142
                /note="AluSx repeat: matches 1. .287 of consensus"
repeat_region   44450. .44600
                /note="MIR repeat: matches 106. .262 of consensus"

Alignment Scores:
Pred. No.:                563             Length:         192867
Score:                    57.00           Matches:        10
Percent Similarity:       81.25%          Conservative:   3
Best Local Similarity:    62.50%          Mismatches:     3
Query Match:              57.58%          Indels:         0
DB:                       9               Gaps:           0

US-10-799-747-116 (1-20) x AL358815 (1-192867)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ::::::||||||   |||   ||||||   ||||||||||||:::|||
Db     183796 TCTAGTCATTCCCATCTTTGCTTTCTGAGCTGGACTCCTTACAGTCTG 183749

RESULT 12
AC024680/c
LOCUS       AC024680               81715 bp    DNA     linear   PRI 25-JUN-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-347D13, complete sequence.
ACCESSION   AC024680
VERSION     AC024680.10  GI:21553261
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 81715)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-347D13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 81715)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
```

```
JOURNAL     Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 81715)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 81715)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 25, 2002 this sequence version replaced gi:21536046.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7498
                Center clone name: 347_D_13
            --------------


            Only the last 81.7 kb of this clone are being submitted.
            The remainder overlaps accession number AC079089 [WICGR project
            L10641].
FEATURES             Location/Qualifiers
     source          1. .81715
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-347D13"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(1. .37)
                     /rpt_family="Alu"
     repeat_region   complement(40. .342)
                     /rpt_family="AluSq"
     repeat_region   complement(810. .867)
                     /rpt_family="MER5B"
     repeat_region   885. .1262
                     /rpt_family="AluY"
     repeat_region   1753. .2182
                     /rpt_family="MLT1D"
     repeat_region   3424. .3730
                     /rpt_family="AluSx"
     repeat_region   3735. .3954
                     /rpt_family="MIR"
     repeat_region   complement(4239. .4482)
                     /rpt_family="MIR"
     repeat_region   4892. .5177
                     /rpt_family="AluJb"
     repeat_region   5183. .5205
                     /rpt_family="AT_rich"
     repeat_region   complement(5327. .5467)
                     /rpt_family="MIR"
     repeat_region   complement(5667. .5997)
                     /rpt_family="AluSq"
     repeat_region   complement(6029. .6085)
                     /rpt_family="L2"
     repeat_region   complement(6946. .7228)
                     /rpt_family="AluSg"
     repeat_region   8128. .8188
                     /rpt_family="AT_rich"
     repeat_region   complement(9341. .9641)
                     /rpt_family="AluSq"
     repeat_region   10146. .10272
                     /rpt_family="LTR16C"
     repeat_region   complement(11117. .11421)
                     /rpt_family="AluJb"
     repeat_region   11524. .11827
                     /rpt_family="AluSx"
     repeat_region   complement(12446. .12883)
                     /rpt_family="L2"
     repeat_region   13082. .13697
                     /rpt_family="L2"
     repeat_region   complement(13721. .14031)
                     /rpt_family="AluSx"
     repeat_region   complement(14457. .14642)
                     /rpt_family="MER5A"
     repeat_region   15672. .16005
                     /rpt_family="MER2"
     repeat_region   complement(16177. .16636)
                     /rpt_family="MER74B"
     repeat_region   17532. .17642
                     /rpt_family="L1MB5"
     repeat_region   complement(18676. .18979)
                     /rpt_family="L2"
     repeat_region   19877. .19998
                     /rpt_family="FRAM"
     repeat_region   20000. .20022
                     /rpt_family="(CAAA)n"
     repeat_region   20271. .20579
                     /rpt_family="AluSg"
     repeat_region   complement(21140. .21243)
                     /rpt_family="L1MC4"
     repeat_region   complement(21275. .21861)
                     /rpt_family="L1MC4"
     repeat_region   21862. .22486
```

```
                     /rpt_family="MER82"
     repeat_region   complement(22487. .22530)
                     /rpt_family="L1MC4"
     repeat_region   complement(22537. .22845)
                     /rpt_family="L1MC4a"
     repeat_region   22859. .22911
                     /rpt_family="(TCTA)n"
     repeat_region   complement(22913. .23034)
                     /rpt_family="L1MC4"
     repeat_region   23402. .23422
                     /rpt_family="AT_rich"
     repeat_region   23642. .23725
                     /rpt_family="AT_rich"
     repeat_region   complement(24308. .24493)
                     /rpt_family="L1ME"
     repeat_region   26614. .26776
                     /rpt_family="MER103"
     repeat_region   complement(27523. .28419)

Alignment Scores:
Pred. No.:                323             Length:         81715
Score:                    56.00           Matches:        9
Percent Similarity:       78.57%          Conservative:   2
Best Local Similarity:    64.29%          Mismatches:     3
Query Match:              56.57%          Indels:         0
DB:                       9               Gaps:           0

US-10-799-747-116 (1-20) x AC024680 (1-81715)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrPro 14
              |||      ||||||:::|||:::||||||   |||||||||
Db      51079 ATGCCACAACATTCGATTTTGACGTTTTTATTCTGGACCCCA 51038

RESULT 13
AC091047/c
LOCUS       AC091047               165127 bp    DNA     linear   PRI 20-SEP-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-102F4, complete sequence.
ACCESSION   AC091047
VERSION     AC091047.10  GI:23238050
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165127)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-102F4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165127)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 165127)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 165127)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 20, 2002 this sequence version replaced gi:22325339.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10484
                Center clone name: 102_F_4
            --------------.
FEATURES              Location/Qualifiers
     source           1. .165127
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="8"
                      /map="8"
                      /clone="RP11-102F4"
                      /clone_lib="RPCI-11 Human Male BAC"
     repeat_region    66. .176
                      /rpt_family="L1MB5"
     repeat_region    complement(1210. .1513)
                      /rpt_family="L2"
     unsure           1797. .1803
                      /note="<30 qual SNGL region"
     unsure           1821. .1825
```

```
                     /note="<30 qual SNGL region"
     repeat_region   2411. .2532
                     /rpt_family="FRAM"
     repeat_region   2534. .2556
                     /rpt_family="(CAAA)n"
     repeat_region   2805. .3112
                     /rpt_family="AluSg"
     repeat_region   complement(3673. .3776)
                     /rpt_family="L1MC4"
     repeat_region   complement(3808. .4394)
                     /rpt_family="L1MC4"
     repeat_region   4395. .5019
                     /rpt_family="MER82"
     repeat_region   complement(5020. .5063)
                     /rpt_family="L1MC4"
     repeat_region   complement(5070. .5378)
                     /rpt_family="L1MC4a"
     repeat_region   5392. .5444
                     /rpt_family="(TCTA)n"
     repeat_region   complement(5446. .5567)
                     /rpt_family="L1MC4"
     repeat_region   5935. .5955
                     /rpt_family="AT_rich"
     repeat_region   6175. .6258
                     /rpt_family="AT_rich"
     repeat_region   complement(6841. .7026)
                     /rpt_family="L1ME"
     repeat_region   9147. .9309
                     /rpt_family="MER103"
     repeat_region   complement(10056. .10952)
                     /rpt_family="L1PA10"
     repeat_region   complement(12015. .12381)
                     /rpt_family="THE1C"
     repeat_region   12402. .12570
                     /rpt_family="MIR"
     repeat_region   complement(12765. .12978)
                     /rpt_family="MIR"
     repeat_region   complement(13005. .13247)
                     /rpt_family="AluSc"
     repeat_region   complement(13285. .13325)
                     /rpt_family="AluSc"
     repeat_region   complement(14540. .14604)
                     /rpt_family="MADE1"
     repeat_region   14605. .14902
                     /rpt_family="AluSx"
     repeat_region   complement(14903. .14928)
                     /rpt_family="MADE1"
     repeat_region   15710. .15892
                     /rpt_family="MIR"
     repeat_region   complement(16294. .16431)
                     /rpt_family="L1MC5"
     repeat_region   complement(16432. .16740)
                     /rpt_family="AluY"
     repeat_region   complement(16741. .17889)
                     /rpt_family="L1MC5"
     repeat_region   complement(18258. .18416)
                     /rpt_family="MIR"
     repeat_region   18631. .18692
                     /rpt_family="(TCCA)n"
     repeat_region   complement(18695. .18804)
                     /rpt_family="L2"
     repeat_region   18867. .18939
                     /rpt_family="MIR"
     repeat_region   19104. .19407
                     /rpt_family="AluSx"
     repeat_region   20410. .20443
                     /rpt_family="(CATA)n"
     repeat_region   20646. .20850
                     /rpt_family="MIR"
     repeat_region   complement(21039. .21092)
                     /rpt_family="MIR"
     repeat_region   21832. .22130
                     /rpt_family="AluSx"
     repeat_region   22358. .22661
                     /rpt_family="AluJo"
     repeat_region   23062. .23127
                     /rpt_family="(CA)n"
     repeat_region   23297. .23425
                     /rpt_family="MIR"
     repeat_region   23516. .23604
                     /rpt_family="MIR"
     repeat_region   complement(23704. .24014)
                     /rpt_family="AluSx"
     repeat_region   24368. .24523
                     /rpt_family="MIR3"
     repeat_region   complement(24564. .24647)
                     /rpt_family="MLT1J"
     repeat_region   complement(24787. .24892)
                     /rpt_family="MLT1J"

Alignment Scores:
Pred. No.:                717             Length:          165127
Score:                    56.00           Matches:         9
Percent Similarity:       78.57%          Conservative:    2
Best Local Similarity:    64.29%          Mismatches:      3
Query Match:              56.57%          Indels:          0
DB:                       9               Gaps:            0

US-10-799-747-116 (1-20) x AC091047 (1-165127)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrPro 14
              |||      ||||||:::|||:::||||||   |||||||||
Db      33611 ATGCCACAACATTCGATTTTGACGTTTTTATTCTGGACCCCA 33570

RESULT 14
AC140381
LOCUS       AC140381               233717 bp    DNA     linear   HTG 23-FEB-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-176M4, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC140381
VERSION     AC140381.1  GI:28475640
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 233717)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 233717)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0176M04
            ----------
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 226925 bases at least Q40
            Consensus quality: 228328 bases at least Q30
            Consensus quality: 229211 bases at least Q20
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
```

```
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1463: contig of 1463 bp in length
            *     1464     1563: gap of unknown length
            *     1564     3655: contig of 2092 bp in length
            *     3656     3755: gap of unknown length
            *     3756     7668: contig of 3913 bp in length
            *     7669     7768: gap of unknown length
            *     7769    11494: contig of 3726 bp in length
            *    11495    11594: gap of unknown length
            *    11595    13492: contig of 1898 bp in length
            *    13493    13592: gap of unknown length
            *    13593    17610: contig of 4018 bp in length
            *    17611    17710: gap of unknown length
            *    17711    23210: contig of 5500 bp in length
            *    23211    23310: gap of unknown length
            *    23311    27592: contig of 4282 bp in length
            *    27593    27692: gap of unknown length
            *    27693    34114: contig of 6422 bp in length
            *    34115    34214: gap of unknown length
            *    34215    45449: contig of 11235 bp in length
            *    45450    45549: gap of unknown length
            *    45550    56906: contig of 11357 bp in length
            *    56907    57006: gap of unknown length
            *    57007    67070: contig of 10064 bp in length
            *    67071    67170: gap of unknown length
            *    67171    80062: contig of 12892 bp in length
            *    80063    80162: gap of unknown length
            *    80163   104523: contig of 24361 bp in length
            *   104524   104623: gap of unknown length
            *   104624   126747: contig of 22124 bp in length
            *   126748   126847: gap of unknown length
            *   126848   156641: contig of 29794 bp in length
            *   156642   156741: gap of unknown length
            *   156742   188299: contig of 31558 bp in length
            *   188300   188399: gap of unknown length
            *   188400   233717: contig of 45318 bp in length.
FEATURES             Location/Qualifiers
     source          1. .233717
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-176M4"
     misc_feature    1. .1463
                     /note="assembly_name:Contig27"
     misc_feature    1564. .3655
                     /note="assembly_name:Contig28"
     misc_feature    3756. .7668
                     /note="assembly_name:Contig30"
     misc_feature    7769. .11494
                     /note="assembly_name:Contig31"
     misc_feature    11595. .13492
                     /note="assembly_name:Contig32"
     misc_feature    13593. .17610
                     /note="assembly_name:Contig33"
     misc_feature    17711. .23210
                     /note="assembly_name:Contig34"
     misc_feature    23311. .27592
                     /note="assembly_name:Contig35"
     misc_feature    27693. .34114
                     /note="assembly_name:Contig36"
     misc_feature    34215. .45449
                     /note="assembly_name:Contig37"
     misc_feature    45550. .56906
                     /note="assembly_name:Contig38"
     misc_feature    57007. .67070
                     /note="assembly_name:Contig39"
     misc_feature    67171. .80062
                     /note="assembly_name:Contig40"
     misc_feature    80163. .104523
                     /note="assembly_name:Contig41"
     misc_feature    104624. .126747
                     /note="assembly_name:Contig42"
     misc_feature    126848. .156641
                     /note="assembly_name:Contig43"
     misc_feature    156742. .188299
                     /note="assembly_name:Contig44"
     misc_feature    188400. .233717
                     /note="assembly_name:Contig45"
ORIGIN

Alignment Scores:
Pred. No.:                1.06e+03        Length:          233717
Score:                    56.00           Matches:         9
Percent Similarity:       77.78%          Conservative:    5
Best Local Similarity:    50.00%          Mismatches:      4
Query Match:              56.57%          Indels:          0
DB:                       2               Gaps:            0

US-10-799-747-116 (1-20) x AC140381 (1-233717)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              :::|||||||||:::|||         :::|||   ||||||:::|||:::|||
Db     126482 TCTGCCCACTCTCTGCTTGTTCTCGCCATTTGGAGACCCTACAGTCTGCGTTCN 126535

RESULT 15
AL137138/c
LOCUS       AL137138               98558 bp    DNA     linear   PRI 31-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-117I13 on chromosome 13,
            complete sequence.
ACCESSION   AL137138
VERSION     AL137138.13  GI:14280387
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98558)
  AUTHORS   Tromans,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 3, 2001 this sequence version replaced gi:13785011.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-117I13 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
```

```
          RP11-117I13 It may be shorter because we sequence overlapping
          sections only once, except for a 100 base overlap.
          The true left end of clone RP11-117I13 is at 1 in this sequence.
          The true left end of clone RP11-366O17 is at 98459 in this
          sequence.
FEATURES             Location/Qualifiers
     source          1. .98558
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-117I13"
                     /clone_lib="RPCI-11.1"
     repeat_region   1380. .1587
                     /note="MER58A repeat: matches 6. .224 of consensus"
     repeat_region   4303. .4506
                     /note="L1MA9 repeat: matches 6070. .6270 of consensus"
     repeat_region   4536. .4844
                     /note="AluSg repeat: matches 1. .305 of consensus"
     repeat_region   5066. .5362
                     /note="AluSx repeat: matches 1. .298 of consensus"
     repeat_region   5544. .5674
                     /note="L2 repeat: matches 2619. .2748 of consensus"
     repeat_region   7241. .8826
                     /note="L1PB1 repeat: matches 4551. .6148 of consensus"
     repeat_region   8828. .9507
                     /note="L1M1 repeat: matches 2438. .2734 of consensus"
     repeat_region   9596. .10327
                     /note="L1PBa repeat: matches -237. .1442 of consensus"
     repeat_region   10350. .11388
                     /note="L1PBa repeat: matches -1537. .-412 of consensus"
     misc_feature    12251. .12425
                     /note="match: STS: Em:G15057"
     repeat_region   12414. .12652
                     /note="MER57-internal repeat: matches 7155. .7387 of
                     consensus"
     repeat_region   12876. .13029
                     /note="MLT1H repeat: matches 178. .340 of consensus"
     repeat_region   13220. .13577
                     /note="MLT1I repeat: matches 43. .410 of consensus"
     repeat_region   13578. .13658
                     /note="L2 repeat: matches 2420. .2496 of consensus"
     repeat_region   13659. .14025
                     /note="THE1C repeat: matches 1. .369 of consensus"
     repeat_region   14026. .14254
                     /note="L2 repeat: matches 2201. .2420 of consensus"
     repeat_region   14255. .14532
                     /note="AluJo repeat: matches 1. .280 of consensus"
     repeat_region   14533. .14844
                     /note="L2 repeat: matches 1862. .2201 of consensus"
     repeat_region   15852. .16106
                     /note="L2 repeat: matches 2201. .2456 of consensus"
     repeat_region   16209. .16274
                     /note="L2 repeat: matches 2133. .2202 of consensus"
     repeat_region   16270. .16309
                     /note="L2 repeat: matches 2464. .2503 of consensus"
     repeat_region   16351. .16819
                     /note="MER41B repeat: matches 130. .596 of consensus"
     repeat_region   17215. .17470
                     /note="L2 repeat: matches 2438. .2701 of consensus"
     repeat_region   17722. .17851
                     /note="MIR repeat: matches 61. .191 of consensus"
     repeat_region   18490. .18521
                     /note="16 copies 2 mer tg 93% conserved"
     repeat_region   19327. .19402
                     /note="MLT1F repeat: matches 89. .161 of consensus"
     repeat_region   19430. .19465
                     /note="18 copies 2 mer tc 86% conserved"
     repeat_region   19536. .19782
                     /note="MLT1F repeat: matches 244. .546 of consensus"
     repeat_region   20422. .20618
                     /note="L2 repeat: matches 2554. .2747 of consensus"
     repeat_region   21524. .21645
                     /note="AluSp/q repeat: matches 187. .303 of consensus"
     repeat_region   21647. .21868
                     /note="AluSx repeat: matches 1. .222 of consensus"
     repeat_region   22326. .22377
                     /note="MIR repeat: matches 47. .97 of consensus"
     repeat_region   24000. .24027
                     /note="14 copies 2 mer tg 89% conserved"
     repeat_region   25338. .25379
                     /note="MIR repeat: matches 58. .99 of consensus"
     repeat_region   25652. .25812
                     /note="MER44A repeat: matches 5. .170 of consensus"
     repeat_region   25813. .26530
                     /note="Tigger3(Golem) repeat: matches 2277. .3028 of
                     consensus"
     repeat_region   26524. .26594
                     /note="Tigger3(Golem) repeat: matches 36. .106 of
                     consensus"
     repeat_region   26655. .26806
                     /note="MER44A repeat: matches 166. .329 of consensus"
     repeat_region   26817. .27149
                     /note="MER46C repeat: matches 3. .338 of consensus"
     repeat_region   27514. .28149
                     /note="LTR5 repeat: matches 1. .623 of consensus"
     misc_feature    complement(27967. .28529)
                     /note="match: STS: Em:G42855"
     repeat_region   28229. .28529
                     /note="LTR5 repeat: matches 675. .969 of consensus"
     repeat_region   29151. .29504
                     /note="THE1A repeat: matches 1. .354 of consensus"
     repeat_region   29676. .29814
                     /note="L2 repeat: matches 2597. .2750 of consensus"
     repeat_region   31144. .31774
                     /note="L1ME1 repeat: matches 5460. .6156 of consensus"
     repeat_region   31775. .31911
                     /note="AluJo repeat: matches 1. .85 of consensus"
     repeat_region   31912. .32186
                     /note="AluSc repeat: matches 1. .281 of consensus"
     repeat_region   32187. .32371
                     /note="AluJo repeat: matches 85. .288 of consensus"
     repeat_region   32372. .32588
                     /note="L1ME1 repeat: matches 5254. .5460 of consensus"
     repeat_region   33053. .33329
                     /note="AluSq repeat: matches 1. .276 of consensus"
     repeat_region   33807. .33871
                     /note="L2 repeat: matches 2622. .2687 of consensus"
     repeat_region   34562. .34746
                     /note="MER5A repeat: matches 1. .188 of consensus"
     repeat_region   35475. .35635
                     /note="MIR repeat: matches 78. .230 of consensus"
     repeat_region   35840. .36257
                     /note="L2 repeat: matches 1450. .1892 of consensus"
     repeat_region   38733. .39038
                     /note="MLT1J repeat: matches 142. .468 of consensus"
     repeat_region   39053. .39147
                     /note="MER86 repeat: matches 88. .182 of consensus"
     repeat_region   39567. .39723
                     /note="L1MEc repeat: matches 1818. .1984 of consensus"
     repeat_region   40423. .40607
                     /note="L1M4 repeat: matches 2680. .2886 of consensus"
     repeat_region   40608. .41117
                     /note="LTR40a repeat: matches 1. .519 of consensus"
     repeat_region   41118. .41187
                     /note="L1M4 repeat: matches 2886. .2950 of consensus"
     repeat_region   41302. .41405
                     /note="52 copies 2 mer aa 60% conserved"
     repeat_region   42443. .42709
                     /note="L2 repeat: matches 2457. .2710 of consensus"
     repeat_region   43036. .43484
                     /note="MLT1C repeat: matches 7. .466 of consensus"
     repeat_region   44299. .44582
                     /note="AluSx repeat: matches 1. .296 of consensus"
     repeat_region   46846. .46936
                     /note="MER91A repeat: matches 17. .113 of consensus"
```

```
     repeat_region   48664. .49299
                     /note="MER4D repeat: matches 362. .981 of consensus"
     repeat_region   49298. .49537
                     /note="MER49 repeat: matches 2. .241 of consensus"
     repeat_region   49608. .49961
                     /note="LTR37A repeat: matches 91. .421 of consensus"
     repeat_region   49987. .50012
                     /note="13 copies 2 mer gt 100% conserved"
     repeat_region   51260. .51623
                     /note="THE1B repeat: matches 1. .364 of consensus"

Alignment Scores:
Pred. No.:                607             Length:         98558
Score:                    55.00           Matches:        9
Percent Similarity:       100.00%         Conservative:   5
Best Local Similarity:    64.29%          Mismatches:     0
Query Match:              55.56%          Indels:         0
DB:                       9               Gaps:           0

US-10-799-747-116 (1-20) x AL137138 (1-98558)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyr 15
              |||:::|||:::|||:::||||||||||||:::||||||:::
Db      25528 GCAAGCCACACAGTCATCTCCTTTCTGCTATATACACCTTTC 25487


Search completed: July 21, 2004, 02:13:58
Job time : 1837 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 00:02:30 ; Search time 348 Seconds
(without alignments)
244.149 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 99
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161415_21434/app_query.fasta_1.199
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747_@CGN_1_1_708_@runat_19072004_161415_21434 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	98	99.0	1434	2	AAX37452	Aax37452 Human sec
	2	98	99.0	1434	7	ADA39771	Ada39771 Human sec
	3	98	99.0	1434	7	ACC50424	Acc50424 Human sec
c	4	56	56.6	8417	4	AAK89495	Aak89495 Human dig
	5	56	56.6	8417	4	AAS32681	Aas32681 Human gen
	6	53	53.5	5760	4	AAS32472	Aas32472 Human gen
	7	53	53.5	5763	4	AAS32471	Aas32471 Human gen
	8	52	52.5	2877	7	ABZ36220	Abz36220 Human sec
c	9	52	52.5	36785	4	AAK82208	Aak82208 Human imm
c	10	51	51.5	400	5	AAF64369	Aaf64369 Novel hum
c	11	51	51.5	505	4	AAH03292	Aah03292 Human cDN
c	12	51	51.5	708	5	AAF64473	Aaf64473 Novel hum
	13	51	51.5	1544	2	AAT86839	Aat86839 cDNA enco
c	14	51	51.5	1568	4	AAH15239	Aah15239 Human cDN
c	15	51	51.5	2500	4	AAH77480	Aah77480 Mandrill
c	16	51	51.5	9641	4	AAH77486	Aah77486 Mandrill
c	17	50	50.5	620	7	ABT40513	Abt40513 Toxicity
c	18	50	50.5	620	9	ADB50026	Adb50026 Primary r
	19	49	49.5	6708	9	ADC56730	Adc56730 Thale cre
	20	49	49.5	6708	9	ADC56727	Adc56727 Open read
	21	49	49.5	7149	9	ADC56728	Adc56728 Thale cre
	22	49	49.5	11000	9	ADC56726	Adc56726 Thale cre
c	23	49	49.5	32768	2	AAX13065	Aax13065 Enterococ
c	24	49	49.5	32768	6	ABS98860	Abs98860 Enterococ
c	25	48	48.5	207	7	ABX88609	Abx88609 Corn ear-
c	26	48	48.5	290	2	AAV88159	Aav88159 EST clone
	27	48	48.5	375	4	AAS39268	Aas39268 Novel hum
	28	48	48.5	404	3	AAF11614	Aaf11614 Aspergill
c	29	48	48.5	481	6	ABN64079	Abn64079 Human can
c	30	48	48.5	485	5	ABV45709	Abv45709 Human pro
	31	48	48.5	495	6	ABQ57387	Abq57387 Human col
c	32	48	48.5	612	6	ABT07938	Abt07938 Human lun
c	33	48	48.5	648	5	ABV15807	Abv15807 Human pro
c	34	48	48.5	675	5	ABV45609	Abv45609 Human pro
	35	48	48.5	804	4	AAH32323	Aah32323 Human olf
	36	48	48.5	939	7	ABZ77940	Abz77940 Human G p
	37	48	48.5	940	9	ADC86574	Adc86574 Human GPC
	38	48	48.5	960	4	AAD12959	Aad12959 Human G-p
	39	48	48.5	960	5	AAS42282	Aas42282 Human cDN
	40	48	48.5	960	6	ABZ43079	Abz43079 Human GPC
	41	48	48.5	960	6	ABK68434	Abk68434 Human DNA
	42	48	48.5	960	6	ABK37568	Abk37568 DNA encod
	43	48	48.5	961	6	AAD28710	Aad28710 Human G-p
	44	48	48.5	970	6	ABS58783	Abs58783 Human G-p
	45	48	48.5	970	6	AAD28709	Aad28709 Human G-p

ALIGNMENTS

RESULT 1
AAX37452
ID AAX37452 standard; cDNA; 1434 BP.
XX
AC AAX37452;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein cDNA fragment containing gene 2.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS Homo sapiens.
XX
PN WO9918208-A1.
XX
PD 15-APR-1999.
XX
PF 01-OCT-1998; 98WO-US020775.
XX
PR 02-OCT-1997; 97US-0060833P.
PR 02-OCT-1997; 97US-0060836P.
PR 02-OCT-1997; 97US-0060837P.
PR 02-OCT-1997; 97US-0060838P.
PR 02-OCT-1997; 97US-0060839P.

PR 02-OCT-1997; 97US-0060843P.
PR 02-OCT-1997; 97US-0060862P.
PR 02-OCT-1997; 97US-0060866P.
PR 02-OCT-1997; 97US-0060874P.
PR 02-OCT-1997; 97US-0060880P.
PR 02-OCT-1997; 97US-0060884P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
PI Lafleur DW, Shi Y;
XX
DR WPI; 1999-264022/22.
DR P-PSDB; AAY07853.
XX
PT New isolated human genes and the secreted polypeptides they encode.
XX
PS Claim 1a; Page 228; 368pp; English.
XX
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in AAY07852-Y07993 and the encoding nucleic acids are
CC represented in AAX37451-X37552
XX
SQ Sequence 1434 BP; 480 A; 204 C; 250 G; 495 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.:	7.13e-07	Length:	1434
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AAX37452 (1-1434)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563
```

RESULT 2
ADA39771
ID ADA39771 standard; cDNA; 1434 BP.
XX
AC ADA39771;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human secreted protein encoding cDNA.
XX
KW Human; secreted protein; cancer; hyperproliferative disorder;
KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
KW anaemia; allergic reaction; asthma; cardiovascular disorder;
KW wound healing; cytostatic; immunosuppressive; nootropic; neuroprotective;
KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
KW vulnerary; cardiant; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO2002102993-A2.
XX
PD 27-DEC-2002.
XX
PF 19-MAR-2002; 2002WO-US008123.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-175238/17.
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating cancer or other hyperproliferative disorder,
PT asthma, allergies or AIDS.
XX
PS Claim 9; SEQ ID NO 153; 3205pp; English.
XX
CC The invention relates to novel genes ADA39629-ADA40565 and proteins
CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.:	7.13e-07	Length:	1434
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0
DB:	7	Gaps:	0

US-10-799-747-116 (1-20) x ADA39771 (1-1434)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563
```

RESULT 3
ACC50424
ID ACC50424 standard; cDNA; 1434 BP.

XX
AC ACC50424;
XX
DT 12-JUN-2003 (first entry)
XX
DE Human secreted protein coding sequence, SEQ ID 91.
XX
KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
KW vulnerary; antiinflammatory; nootropic; neuroprotective;
KW antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200295010-A2.
XX
PD 28-NOV-2002.
XX
PF 19-MAR-2002; 2002WO-US009785.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-129429/12.
XX
PT Novel human secreted proteins, useful for detecting, preventing,
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
PT disorders such as arrhythmia.
XX
PS Claim 21; SEQ ID NO 91; 1881pp; English.
XX
CC The present invention relates to novel human secreted proteins (ABR47633-
CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
CC and their coding sequences are useful for the preparation of a diagnostic
CC or pharmaceutical composition for diagnosing or treating a cardiovascular
CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
CC arteriosclerosis and myocardial ischaemia), neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, renal disorders,
CC proliferative disorders and/or cancerous diseases and conditions, for
CC wound healing and epithelial cell proliferation, to treat inflammation or
CC infection, for treating thrombosis and arteriosclerosis, for treating or
CC preventing neural damage which occurs in neuronal disorders or
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
CC disease, to enhance bone and periodontal regeneration and aid in tissue
CC transplants or bone grafts, to prevent skin aging or hair loss, to
CC stimulate growth and differentiation of haematopoietic cells and bone
CC marrow cells when used in combination with other cytokines, to maintain
CC organs before transplantation or for supporting cell culture of primary
CC tissues, to increase or decrease differentiation or proliferation of
CC embryonic stem cells, or to modulate mammalian characteristics or
CC metabolism. Note: The sequence data for this patent was published in
CC electronic format and is available from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.:	7.13e-07	Length:	1434
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0
DB:	7	Gaps:	0

US-10-799-747-116 (1-20) x ACC50424 (1-1434)

Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563

RESULT 4
AAK89495/c
ID AAK89495 standard; DNA; 8417 BP.
XX
AC AAK89495;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 3071.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001324.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Disclosure; SEQ ID NO 3071; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 8417 BP; 2708 A; 1429 C; 1657 G; 2621 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.:	102	Length:	8417
Score:	56.00	Matches:	9
Percent Similarity:	78.57%	Conservative:	2
Best Local Similarity:	64.29%	Mismatches:	3
Query Match:	56.57%	Indels:	0
DB:	4	Gaps:	0

US-10-799-747-116 (1-20) x AAK89495 (1-8417)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrPro 14
              |||      ||||||:::|||:::||||||   |||||||||
Db       7484 ATGCCACAACATTCGATTTTGACGTTTTTATTCTGGACCCCA 7443
```

RESULT 5
AAS32681
ID AAS32681 standard; DNA; 8417 BP.
XX
AC AAS32681;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 635.
XX
KW Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS Homo sapiens.
XX
PN WO200155319-A2.
XX
PD 02-AUG-2001.

XX
PF 17-JAN-2001; 2001WO-US001335.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457726/49.
XX

PT Isolated polypeptide for treating, preventing and prognosing disorders
PT related to the endocrine system including endocrine disorders,
PT reproductive disorders, and gastrointestinal disorders and also for
PT testing and detection e.g. diagnosis.
XX
PS Disclosure; SEQ ID NO 635; 558pp; English.
XX
CC The invention relates to cDNAs encoding novel human endocrine antigens or
CC a fragment having biological activity, a domain, an epitope, full length
CC protein, variant, allelic variant or a species homologue of the
CC cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition when administered (e.g. by
CC gene therapy or antisense-therapy). Identifying mutations in the genes
CC coding for the antigens is useful for diagnosing a pathological condition
CC or a susceptibility to a pathological condition. The DNAs, antigens and
CC antibodies raised against the antigens useful for treating, preventing
CC and/ or prognosing disorders related to the endocrine system or hormone
CC imbalance or reproductive disorders, cancers of endocrine tissues,
CC disorders of the pancreas (e.g. diabetes mellitus), the adrenal glands
CC (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC of diseases and disorders are given in the specification. The present
CC sequence is genomic DNA fragment form a gene encoding an endocrine
CC antigen of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 8417 BP; 2621 A; 1657 C; 1429 G; 2708 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.:	102	Length:	8417
Score:	56.00	Matches:	9
Percent Similarity:	78.57%	Conservative:	2
Best Local Similarity:	64.29%	Mismatches:	3
Query Match:	56.57%	Indels:	0
DB:	4	Gaps:	0

US-10-799-747-116 (1-20) x AAS32681 (1-8417)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrPro 14
              |||      ||||||:::|||:::||||||   |||||||||
Db        934 ATGCCACAACATTCGATTTTGACGTTTTTATTCTGGACCCCA 975
```

RESULT 6
AAS32472
ID AAS32472 standard; DNA; 5760 BP.
XX
AC AAS32472;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 426.
XX
KW Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS Homo sapiens.
XX
PN WO200155319-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001335.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457726/49.
XX
PT Isolated polypeptide for treating, preventing and prognosing disorders
PT related to the endocrine system including endocrine disorders,
PT reproductive disorders, and gastrointestinal disorders and also for
PT testing and detection e.g. diagnosis.
XX
PS Disclosure; SEQ ID NO 426; 558pp; English.
XX
CC The invention relates to cDNAs encoding novel human endocrine antigens or
CC a fragment having biological activity, a domain, an epitope, full length
CC protein, variant, allelic variant or a species homologue of the
CC cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition when administered (e.g. by
CC gene therapy or antisense-therapy). Identifying mutations in the genes
CC coding for the antigens is useful for diagnosing a pathological condition
CC or a susceptibility to a pathological condition. The DNAs, antigens and
CC antibodies raised against the antigens useful for treating, preventing
CC and/ or prognosing disorders related to the endocrine system or hormone
CC imbalance or reproductive disorders, cancers of endocrine tissues,
CC disorders of the pancreas (e.g. diabetes mellitus), the adrenal glands
CC (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC of diseases and disorders are given in the specification. The present
CC sequence is genomic DNA fragment form a gene encoding an endocrine
CC antigen of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 5760 BP; 1725 A; 1199 C; 1068 G; 1768 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	208	Length:	5760
Score:	53.00	Matches:	10
Percent Similarity:	68.75%	Conservative:	1
Best Local Similarity:	62.50%	Mismatches:	5
Query Match:	53.54%	Indels:	0
DB:	4	Gaps:	0

US-10-799-747-116 (1-20) x AAS32472 (1-5760)

```
Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||   ||||||   |||||||||   |||||       |||:::|||
Db        567 GCCCGGCACTCAACCCTGAGCTTTTGCCTTTGGGTTGTATATTCTCTT 614
```

RESULT 7
AAS32471
ID AAS32471 standard; DNA; 5763 BP.
XX
AC AAS32471;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 425.
XX
KW Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS Homo sapiens.
XX
PN WO200155319-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001335.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457726/49.
XX
PT Isolated polypeptide for treating, preventing and prognosing disorders
PT related to the endocrine system including endocrine disorders,
PT reproductive disorders, and gastrointestinal disorders and also for
PT testing and detection e.g. diagnosis.
XX
PS Disclosure; SEQ ID NO 425; 558pp; English.
XX
CC The invention relates to cDNAs encoding novel human endocrine antigens or
CC a fragment having biological activity, a domain, an epitope, full length
CC protein, variant, allelic variant or a species homologue of the
CC cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition when administered (e.g. by
CC gene therapy or antisense-therapy). Identifying mutations in the genes
CC coding for the antigens is useful for diagnosing a pathological condition
CC or a susceptibility to a pathological condition. The DNAs, antigens and
CC antibodies raised against the antigens useful for treating, preventing

```
CC   and/ or prognosing disorders related to the endocrine system or hormone
CC   imbalance or reproductive disorders, cancers of endocrine tissues,
CC   disorders of the pancreas (e.g. diabetes mellitus), the adrenal glands
CC   (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC   hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC   of diseases and disorders are given in the specification. The present
CC   sequence is genomic DNA fragment form a gene encoding an endocrine
CC   antigen of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5763 BP; 1725 A; 1203 C; 1066 G; 1768 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:                209             Length:         5763
Score:                    53.00           Matches:        10
Percent Similarity:       68.75%          Conservative:   1
Best Local Similarity:    62.50%          Mismatches:     5
Query Match:              53.54%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x AAS32471 (1-5763)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||   ||||||   |||||||||   ||||||      |||:::|||
Db        567 GCCCGGCACTCAACCCTGAGCTTTTGCCTTTGGGTTGTATATTCTCTT 614

RESULT 8
ABZ36220
ID   ABZ36220 standard; cDNA; 2877 BP.
XX
AC   ABZ36220;
XX
DT   10-FEB-2003  (first entry)
XX
DE   Human secretory polynucleotide SPTM SEQ ID NO 384.
XX
KW   Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;
KW   asthma; Crohn's disease; neurological disorder; epilepsy; cancer;
KW   Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;
KW   multiple sclerosis; Parkinson's disease; cell proliferative disorder;
KW   anti-inflammatory; immunosuppressive; neuroprotective; nootropic;
KW   neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;
KW   antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;
KW   secretory polynucleotide; secretory protein; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200283876-A2.
XX
PD   24-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-US009921.
XX
PR   29-MAR-2001; 2001US-0280067P.
PR   29-MAR-2001; 2001US-0280068P.
PR   16-MAY-2001; 2001US-0291280P.
PR   17-MAY-2001; 2001US-0291829P.
PR   17-MAY-2001; 2001US-0291849P.
PR   19-JUN-2001; 2001US-0299428P.
PR   20-JUN-2001; 2001US-0299776P.
PR   20-JUN-2001; 2001US-0300001P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Daffo A,  Jones AL,  Tran AB,  Dahl CR,  Gietzen D,  Chinn J;
PI   Dufour GE,  Hillman JL,  Yu JY,  Tuason O,  Yap PE,  Amshey SR;
PI   Daughtery SC,  Dam TC,  Liu TF,  Nguyen DA,  Kleefeld Y,  Gerstin EH;
PI   Peralta CH,  David MH,  Lewis SA,  Chen AJ,  Panzer SR,  Harris B;
PI   Flores V,  Marwaha R,  Lo A,  Lan RY,  Urashka ME;
XX
DR   WPI; 2003-075543/07.
DR   P-PSDB; ABP75778.
XX
PT   New human secretory proteins and polynucleotides, useful for diagnosing,
PT   treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),
PT   neurological disorders (e.g. Alzheimer's), or cell proliferations or
PT   cancers.
XX
PS   Claim 1; SEQ ID NO 384; 458pp + Sequence Listing; English.
XX
CC   The invention relates to a secretory polynucleotide (designated sptm)
CC   comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a
CC   naturally occurring polynucleotide sequence at least 90 % identical to
CC   the polynucleotide sequence, a polynucleotide complementary to them or an
CC   RNA equivalent of them. The polypeptide or polynucleotide are useful for
CC   treating, preventing or diagnosing a disease or condition associated with
CC   the expression of functional SPTM. These are particularly useful for
CC   diagnosing, treating or preventing autoimmune/inflammatory disorders
CC   (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
CC   disease), neurological disorders (e.g. epilepsy, Huntington's disease,
CC   dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
CC   multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
CC   schizophrenia or amnesia), or cell proliferative disorders (e.g.
CC   psoriasis, polycythemia vera, or cancers including adenocarcinoma,
CC   leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
CC   breast, cervix or prostate). Note: The sequence data for this patent did
CC   not form part of the printed specfication, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2877 BP; 854 A; 609 C; 504 G; 910 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                131             Length:         2877
Score:                    52.00           Matches:        9
Percent Similarity:       73.33%          Conservative:   2
Best Local Similarity:    60.00%          Mismatches:     4
Query Match:              52.53%          Indels:         0
DB:                       7               Gaps:           0

US-10-799-747-116 (1-20) x ABZ36220 (1-2877)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAla 16
              |||   ||||||   |||||||||:::::|||   |||   |||
Db       2193 GCTCCTCACAGCAAATTATCATTCATTATTTGGGTGCCACAGGCC 2237

RESULT 9
AAK82208/c
ID   AAK82208 standard; DNA; 36785 BP.
XX
AC   AAK82208;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37020.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
```

PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 37020; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 36785 BP; 9938 A; 8156 C; 8241 G; 10450 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.04e+03	Length:	36785
Score:	52.00	Matches:	9
Percent Similarity:	73.33%	Conservative:	2
Best Local Similarity:	60.00%	Mismatches:	4
Query Match:	52.53%	Indels:	0
DB:	4	Gaps:	0

US-10-799-747-116 (1-20) x AAK82208 (1-36785)

```
Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAla 16
              |||   ||||||   |||||||||:::::|||   |||   |||
Db      13072 GCTCCTCACAGCAAATTATCATTCATTATTTGGGTGCCACAGGCC 13028
```

RESULT 10
AAF64369/c
ID AAF64369 standard; cDNA; 400 BP.
XX
AC AAF64369;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 125.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US018374.
XX
PR 02-JUL-1999; 99US-0142310P.
PR 02-JUL-1999; 99US-0142311P.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;
PI Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;
PI Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;
PI Leshkowitiz D, Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences.
XX
PS Claim 9; Page 562; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia
XX
SQ Sequence 400 BP; 123 A; 64 C; 119 G; 94 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	17.2	Length:	400
Score:	51.00	Matches:	11
Percent Similarity:	70.00%	Conservative:	3
Best Local Similarity:	55.00%	Mismatches:	4
Query Match:	51.52%	Indels:	2
DB:	5	Gaps:	1

US-10-799-747-116 (1-20) x AAF64369 (1-400)

```
Qy          2 AlaAlaHisSerValLeuSer------PheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              |||:::||||||       |||      ||||||:::|||||||||   ||||||   :::
Db        198 GCCTCCCACTCTTGTGCAAGTGATCACTTCCTCATTTGGACGCCCATTGCACTCTCTGCA 139
```

RESULT 11
AAH03292/c
ID AAH03292 standard; cDNA; 505 BP.
XX
AC AAH03292;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:127.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 1; SEQ ID NO 127; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the

```
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 505 BP; 167 A; 76 C; 125 G; 134 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.:                22.9          Length:          505
Score:                    51.00         Matches:         11
Percent Similarity:       70.00%        Conservative:    3
Best Local Similarity:    55.00%        Mismatches:      4
Query Match:              51.52%        Indels:          2
DB:                       4             Gaps:            1

US-10-799-747-116 (1-20) x AAH03292 (1-505)

Qy          2 AlaAlaHisSerValLeuSer------PheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              |||:::||||||       |||      ||||||:::|||||||||   ||||||   :::
Db        132 GCCTCCCACTCTTGTGCAAGTGATCACTTCCTCATTTGGACGCCCATTGCACTCTCTGCA 73

RESULT 12
AAF64473/c
ID   AAF64473 standard; cDNA; 708 BP.
XX
AC   AAF64473;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 229.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US018374.
XX
PR   02-JUL-1999;   99US-0142310P.
PR   02-JUL-1999;   99US-0142311P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;
PI   Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;
PI   Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;
PI   Leshkowitiz D, Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences.
XX
PS   Claim 9; Page 577; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia
XX
SQ   Sequence 708 BP; 216 A; 116 C; 182 G; 189 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.:                34.7          Length:          708
Score:                    51.00         Matches:         11
Percent Similarity:       70.00%        Conservative:    3
Best Local Similarity:    55.00%        Mismatches:      4
Query Match:              51.52%        Indels:          2
DB:                       5             Gaps:            1

US-10-799-747-116 (1-20) x AAF64473 (1-708)

Qy          2 AlaAlaHisSerValLeuSer------PheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              |||:::||||||       |||      ||||||:::|||||||||   ||||||   :::
Db        202 GCCTCCCACTCTTGTGCAAGTGATCACTTCCTCATTTGGACGCCCATTGCACTCTCTGCA 143

RESULT 13
AAT86839
ID   AAT86839 standard; cDNA; 1544 BP.
XX
AC   AAT86839;
XX
DT   19-DEC-1997  (first entry)
XX
DE   cDNA encoding rat CC chemokine receptor.
XX
KW   rat; CC chemokine receptor; screen; binding; ligand; ds.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             33. .1100
FT                   /*tag=  a
XX
PN   JP09227599-A.
XX
PD   02-SEP-1997.
XX
PF   22-FEB-1996;   96JP-00035192.
XX
PR   22-FEB-1996;   96JP-00035192.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
DR   WPI; 1997-486426/45.
DR   P-PSDB; AAW29179.
XX
PT   CC chemokine receptor protein - useful to screen for novel binding
PT   compounds.
XX
PS   Claim 4; Page 21-22; 26pp; Japanese.
XX
CC   This cDNA encodes a rat CC chemokine receptor. The receptor can be used
CC   to screen for novel binding compounds and for preparation of antibodies
CC   or antiserum
XX
SQ   Sequence 1544 BP; 415 A; 347 C; 330 G; 452 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                  90.6          Length:          1544
Score:                      51.00         Matches:         10
Percent Similarity:         80.00%        Conservative:    2
Best Local Similarity:      66.67%        Mismatches:      3
Query Match:                51.52%        Indels:          0
DB:                         2             Gaps:            0

US-10-799-747-116 (1-20) x AAT86839 (1-1544)

Qy          3 AlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||   ::::::|||   ||||||||||||||||||||   |||
Db        759 GCTATTACGCTTCTATTCTTCCTCCTCTGGACCCCCTACAATCTG 803

RESULT 14
AAH15239/c
ID   AAH15239 standard; cDNA; 1568 BP.
XX
AC   AAH15239;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:13356.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 13356; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 1568 BP; 447 A; 306 C; 425 G; 390 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                  92.4          Length:          1568
Score:                      51.00         Matches:         11
Percent Similarity:         70.00%        Conservative:    3
Best Local Similarity:      55.00%        Mismatches:      4
Query Match:                51.52%        Indels:          2
DB:                         4             Gaps:            1

US-10-799-747-116 (1-20) x AAH15239 (1-1568)

Qy          2 AlaAlaHisSerValLeuSer------PheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              |||:::|||||       |||       ||||||:::||||||||   ||||||   :::
Db        132 GCCTCCCACTCTTGTGCAAGTGATCACTTCCTCATTTGGACGCCCATTGCACTCTCTGCA 73

RESULT 15
AAH77480/c
ID   AAH77480 standard; DNA; 2500 BP.
XX
AC   AAH77480;
XX
DT   20-NOV-2001  (first entry)
XX
DE   Mandrill immunodeficiency virus SIM27 related coding sequence #4.
XX
KW   Simian immunodeficiency virus; SIV; HIV; gag; env; pol; mandrill;
KW   antibody detection; ds.
XX
OS   Simian immunodeficiency virus.
XX
PN   NO200003889-A.
XX
PD   05-FEB-2001.
XX
PF   28-JUL-2000; 2000NO-00003889.
XX
PR   03-AUG-1999;   99DE-01036003.
XX
PA   (DADE-) DADE BEHRING MARBURG GMBH.
XX
PI   Guertler LG,  Kaptue L,  Zekeng L,  Deloko YBD,  Hauser HP;
XX
DR   WPI; 2001-201061/20.
XX
PT   New mandrill immunodeficiency virus SIM27 useful for detecting antibodies
PT   directed against the virus.
XX
PS   Claim 1; Page 11-12; 47pp; Norwegian.
XX
CC   The present invention relates to a new mandrill immunodeficiency virus
CC   SIM27, which can be used to detect antibodies directed against the virus.
CC   The present sequence is a coding sequence described in the
CC   exemplification of the invention
XX
SQ   Sequence 2500 BP; 872 A; 498 C; 637 G; 483 T; 0 U; 10 Other;

Alignment Scores:
Pred. No.:                  164           Length:          2500
Score:                      51.00         Matches:         8
Percent Similarity:         100.00%       Conservative:    3
Best Local Similarity:      72.73%        Mismatches:      0
Query Match:                51.52%        Indels:          0
DB:                         4             Gaps:            0

US-10-799-747-116 (1-20) x AAH77480 (1-2500)

Qy          7 LeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||||||||:::||||||:::||||||:::|||
```

Db 1453 TTGAGCTTTATACTTTGGTCCCCCTACTCCTTG 1421

Search completed: July 21, 2004, 01:43:18
Job time : 356 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 01:24:25 ; Search time 72 Seconds
(without alignments)
154.153 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 99
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161418_21467/app_query.fasta_1.199
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747_@CGN_1_1_105_@runat_19072004_161418_21467 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	51	51.5	2500	4	US-09-625-972-20	Sequence 20, Appl
c 2	51	51.5	9641	4	US-09-625-972-22	Sequence 22, Appl
3	49	49.5	1368	4	US-09-134-000C-1177	Sequence 1177, Ap
c 4	48	48.5	207	4	US-09-313-294A-7069	Sequence 7069, Ap
c 5	48	48.5	1601	3	US-08-492-459-17	Sequence 17, Appl
c 6	48	48.5	1601	3	US-08-492-459-18	Sequence 18, Appl
c 7	48	48.5	1601	3	US-08-423-752-17	Sequence 17, Appl
c 8	48	48.5	1601	3	US-08-423-752-18	Sequence 18, Appl
c 9	48	48.5	1601	3	US-08-716-873-31	Sequence 31, Appl
c 10	48	48.5	1601	3	US-08-716-873-32	Sequence 32, Appl
c 11	48	48.5	1601	3	US-09-368-431-31	Sequence 31, Appl
c 12	48	48.5	1601	3	US-09-368-431-32	Sequence 32, Appl
c 13	48	48.5	1601	4	US-09-414-006-17	Sequence 17, Appl
c 14	48	48.5	1601	4	US-09-414-006-18	Sequence 18, Appl
c 15	48	48.5	1601	4	US-09-447-223-17	Sequence 17, Appl
c 16	48	48.5	1601	4	US-09-447-223-18	Sequence 18, Appl
17	48	48.5	2385	3	US-08-492-459-1	Sequence 1, Appli
18	48	48.5	2385	3	US-08-492-459-3	Sequence 3, Appli
19	48	48.5	2385	3	US-08-423-752-1	Sequence 1, Appli
20	48	48.5	2385	3	US-08-423-752-3	Sequence 3, Appli
21	48	48.5	2385	3	US-08-716-873-15	Sequence 15, Appl
22	48	48.5	2385	3	US-08-716-873-17	Sequence 17, Appl
23	48	48.5	2385	3	US-09-368-431-15	Sequence 15, Appl
24	48	48.5	2385	3	US-09-368-431-17	Sequence 17, Appl
25	48	48.5	2385	4	US-09-414-006-1	Sequence 1, Appli
26	48	48.5	2385	4	US-09-414-006-3	Sequence 3, Appli
27	48	48.5	2385	4	US-09-447-223-1	Sequence 1, Appli
28	48	48.5	2385	4	US-09-447-223-3	Sequence 3, Appli
c 29	48	48.5	17000	4	US-09-679-299A-18	Sequence 18, Appl
30	47	47.5	2210	4	US-09-016-434-1177	Sequence 1177, Ap
31	46	46.5	292	4	US-09-313-294A-4459	Sequence 4459, Ap
32	46	46.5	1098	4	US-09-718-692-3	Sequence 3, Appli
33	46	46.5	1098	4	US-09-718-852-3	Sequence 3, Appli
34	46	46.5	1098	4	US-09-718-815-3	Sequence 3, Appli
c 35	46	46.5	1296	4	US-09-328-352-467	Sequence 467, App
36	46	46.5	2595	4	US-09-016-434-1178	Sequence 1178, Ap
37	46	46.5	4911	4	US-09-718-692-1	Sequence 1, Appli
38	46	46.5	4911	4	US-09-718-852-1	Sequence 1, Appli
39	46	46.5	4911	4	US-09-718-815-1	Sequence 1, Appli
c 40	45.5	46.0	1230025	4	US-09-198-452A-1	Sequence 1, Appli
41	45	45.5	1495	4	US-09-016-434-1190	Sequence 1190, Ap
42	45	45.5	1495	4	US-09-023-655-1021	Sequence 1021, Ap
c 43	45	45.5	1896	4	US-09-134-000C-616	Sequence 616, App
44	45	45.5	2156	1	US-08-012-988A-1	Sequence 1, Appli
45	45	45.5	2156	4	US-09-023-655-1247	Sequence 1247, Ap

ALIGNMENTS

RESULT 1
US-09-625-972-20/c
; Sequence 20, Application US/09625972
; Patent No. 6566513
; GENERAL INFORMATION:
; APPLICANT: GUERTLER, Lutz Gerhard
; APPLICANT: HAUSER, Hans Peter
; APPLICANT: DONGMO DELOKO, Yvette Beatrice
; APPLICANT: ZEKENG, Leopold
; APPLICANT: KAPTUE, Lazare
; TITLE OF INVENTION: LENTIVIRUS FROM THE GROUP OF IMMUNODEFICIENCY VIRUSES OF DRILL M(
; TITLE OF INVENTION: (MANDRILLUS LEUCOPHAEUS) AND THEIR USE
; FILE REFERENCE: 067595/0106
; CURRENT APPLICATION NUMBER: US/09/625,972
; CURRENT FILING DATE: 2000-07-29
; PRIOR APPLICATION NUMBER: DE 199 36 003.0
; PRIOR FILING DATE: 1999-08-03
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 2500
; TYPE: DNA
; ORGANISM: SIV - viral
US-09-625-972-20

Alignment Scores:

Pred. No.:	36.3	Length:	2500
Score:	51.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	3
Best Local Similarity:	72.73%	Mismatches:	0
Query Match:	51.52%	Indels:	0
DB:	4	Gaps:	0

US-10-799-747-116 (1-20) x US-09-625-972-20 (1-2500)

Qy 7 LeuSerPheLeuLeuTrpThrProTyrAlaLeu 17

```
                  |||||||||:::||||||:::||||||:::|||
Db       1453 TTGAGCTTTATACTTTGGTCCCCCTACTCCTTG 1421

RESULT 2
US-09-625-972-22/c
; Sequence 22, Application US/09625972
; Patent No. 6566513
; GENERAL INFORMATION:
;  APPLICANT: GUERTLER, Lutz Gerhard
;  APPLICANT:  HAUSER, Hans Peter
;  APPLICANT:  DONGMO DELOKO, Yvette Beatrice
;  APPLICANT:  ZEKENG, Leopold
;  APPLICANT:  KAPTUE, Lazare
;  TITLE OF INVENTION: LENTIVIRUS FROM THE GROUP OF IMMUNODEFICIENCY VIRUSES OF DRILL MO
;  TITLE OF INVENTION:  (MANDRILLUS LEUCOPHAEUS) AND THEIR USE
;  FILE REFERENCE: 067595/0106
;  CURRENT APPLICATION NUMBER: US/09/625,972
;  CURRENT FILING DATE:  2000-07-29
;  PRIOR APPLICATION NUMBER: DE 199 36 003.0
;  PRIOR FILING DATE: 1999-08-03
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 9641
;   TYPE: DNA
;   ORGANISM: SIV - viral
US-09-625-972-22

Alignment Scores:
Pred. No.:                189             Length:          9641
Score:                    51.00           Matches:         8
Percent Similarity:       100.00%         Conservative:    3
Best Local Similarity:    72.73%          Mismatches:      0
Query Match:              51.52%          Indels:          0
DB:                       4               Gaps:            0

US-10-799-747-116 (1-20) x US-09-625-972-22 (1-9641)

Qy           7 LeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
                |||||||||:::||||||:::||||||:::|||
Db       1453 TTGAGCTTTATACTTTGGTCCCCCTACTCCTTG 1421

RESULT 3
US-09-134-000C-1177
; Sequence 1177, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1177
;   LENGTH: 1368
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-1177

Alignment Scores:
Pred. No.:                38              Length:          1368
Score:                    49.00           Matches:         9
Percent Similarity:       75.00%          Conservative:    0
Best Local Similarity:    75.00%          Mismatches:      3
Query Match:              49.49%          Indels:          0
DB:                       4               Gaps:            0

US-10-799-747-116 (1-20) x US-09-134-000C-1177 (1-1368)

Qy           4 HisSerValLeuSerPheLeuLeuTrpThrProTyr 15
                |||   |||   |||   ||||||||||||||||||
Db       1022 CACTAAGTGGCGTCATTACTTCTGTGGACTCCTTAC 1057

RESULT 4
US-09-313-294A-7069/c
; Sequence 7069, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 7069
;   LENGTH: 207
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700381075H1
US-09-313-294A-7069

Alignment Scores:
Pred. No.:                5.57            Length:          207
Score:                    48.00           Matches:         9
Percent Similarity:       84.62%          Conservative:    2
Best Local Similarity:    69.23%          Mismatches:      2
Query Match:              48.48%          Indels:          0
DB:                       4               Gaps:            0

US-10-799-747-116 (1-20) x US-09-313-294A-7069 (1-207)

Qy           5 SerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
                :::|||||||  ||||||||||||||||||  :::|||
Db        180 AACGTGCTGGTATTTCTCCTTTGGACCCCACAATCCCTT 142

RESULT 5
US-08-492-459-17/c
; Sequence 17, Application US/08492459
; Patent No. 6015689
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  REGULATION OF AUREOBASIDIN SENSITIVITY IN FUNGUS
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/492,459
;      FILING DATE:  June 20, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,403
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  202-371-8850
;       TELEFAX:
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1601
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  genomic DNA
;     ANTI-SENSE:  Yes
US-08-492-459-17

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       3             Gaps:            0

US-10-799-747-116 (1-20) x US-08-492-459-17 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||       |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 6
US-08-492-459-18/c
; Sequence 18, Application US/08492459
; Patent No. 6015689
;  GENERAL INFORMATION:
;     APPLICANT:  Takashi OKADO et al.
;     TITLE OF INVENTION:  REGULATION OF AUREOBASIDIN SENSITIVITY IN FUNGUS
;     NUMBER OF SEQUENCES:  36
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Wenderoth, Lind & Ponack
;       STREET:  805 Fifteenth Street, N.W., #700
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.A.
;       ZIP:  20005
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 mb
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Wordperfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/492,459
;       FILING DATE:  June 20, 1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/243,403
;       FILING DATE:  May 16, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warren M. Cheek, Jr.
;       REGISTRATION NUMBER:  33,367
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-8850
;       TELEFAX:
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1601
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  mRNA
;     ANTI-SENSE:  Yes
US-08-492-459-18

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       3             Gaps:            0

US-10-799-747-116 (1-20) x US-08-492-459-18 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||       |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 7
US-08-423-752-17/c
; Sequence 17, Application US/08423752
; Patent No. 6022949
;  GENERAL INFORMATION:
;     APPLICANT:  Takashi OKADO et al.
;     TITLE OF INVENTION:  A GENE CODING FOR A PROTEIN REGULATING
;     TITLE OF INVENTION:  AUREOBASIDIN SENSITIVITY
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Wenderoth, Lind & Ponack
;       STREET:  805 Fifteenth Street, N.W., #700
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.A.
;       ZIP:  20005
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Wordperfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/423,752
;       FILING DATE:  April 18, 1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/243,403
;       FILING DATE:  May 16, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warren M. Cheek, Jr.
;       REGISTRATION NUMBER:  33,367
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-8850
;       TELEFAX:
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1601
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  genomic DNA
;     ANTI-SENSE:  Yes
US-08-423-752-17

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       3             Gaps:            0

US-10-799-747-116 (1-20) x US-08-423-752-17 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||       |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802
```

```
RESULT 8
US-08-423-752-18/c
; Sequence 18, Application US/08423752
; Patent No. 6022949
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  A GENE CODING FOR A PROTEIN REGULATING
;    TITLE OF INVENTION:  AUREOBASIDIN SENSITIVITY
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/423,752
;      FILING DATE:  April 18, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,403
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1601
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  mRNA
;    ANTI-SENSE:  Yes
US-08-423-752-18

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       3             Gaps:            0

US-10-799-747-116 (1-20) x US-08-423-752-18 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||      |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 9
US-08-716-873-31/c
; Sequence 31, Application US/08716873
; Patent No. 6194166
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  GENE REGULATING AUREOBASIDIN SENSITIVITY
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,873
;      FILING DATE:  September 20, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1601 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ANTI-SENSE:  Yes
US-08-716-873-31

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       3             Gaps:            0

US-10-799-747-116 (1-20) x US-08-716-873-31 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||      |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 10
US-08-716-873-32/c
; Sequence 32, Application US/08716873
; Patent No. 6194166
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  GENE REGULATING AUREOBASIDIN SENSITIVITY
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,873
;      FILING DATE:  September 20, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1601 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  mRNA
;    ANTI-SENSE:  Yes
US-08-716-873-32

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       3             Gaps:            0

US-10-799-747-116 (1-20) x US-08-716-873-32 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||      |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 11
US-09-368-431-31/c
; Sequence 31, Application US/09368431
; Patent No. 6294651
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  GENE REGULATING AUREOBASIDIN SENSITIVITY
;    TITLE OF INVENTION:  (AS AMENDED)
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  WENDEROTH, LIND & PONACK, L.L.P.
;      STREET:  2033 K Street, N.W., #800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/368,431
;      FILING DATE:  August 5, 1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/716,873
;      FILING DATE:  September 20, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1601 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ANTI-SENSE:  Yes
US-09-368-431-31

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       3             Gaps:            0

US-10-799-747-116 (1-20) x US-09-368-431-31 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||      |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 12
US-09-368-431-32/c
; Sequence 32, Application US/09368431
; Patent No. 6294651
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  GENE REGULATING AUREOBASIDIN SENSITIVITY
;    TITLE OF INVENTION:  (AS AMENDED)
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  WENDEROTH, LIND & PONACK, L.L.P.
;      STREET:  2033 K Street, N.W., #800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/368,431
;      FILING DATE:  August 5, 1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/716,873
;      FILING DATE:  September 20, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1601 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  mRNA
;    ANTI-SENSE:  Yes
US-09-368-431-32

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       3             Gaps:            0
```

```
US-10-799-747-116 (1-20) x US-09-368-431-32 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||       |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 13
US-09-414-006-17/c
; Sequence 17, Application US/09414006
; Patent No. 6348577
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  REGULATION OF AUREOBASIDIN SENSITIVITY (AS
;    TITLE OF INVENTION:  AMENDED)
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack, L.L.P.
;      STREET:  2033 K Street, N.W., #800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/414,006
;      FILING DATE:  October 7, 1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/492,459
;      FILING DATE:  June 20, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,403
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1601
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ANTI-SENSE:  Yes
US-09-414-006-17

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       4             Gaps:            0

US-10-799-747-116 (1-20) x US-09-414-006-17 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||       |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 14
US-09-414-006-18/c
; Sequence 18, Application US/09414006
; Patent No. 6348577
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi OKADO et al.
;    TITLE OF INVENTION:  REGULATION OF AUREOBASIDIN SENSITIVITY (AS
;    TITLE OF INVENTION:  AMENDED)
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack, L.L.P.
;      STREET:  2033 K Street, N.W., #800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/414,006
;      FILING DATE:  October 7, 1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/492,459
;      FILING DATE:  June 20, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,403
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1601
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  mRNA
;    ANTI-SENSE:  Yes
US-09-414-006-18

Alignment Scores:
Pred. No.:                68.2          Length:          1601
Score:                    48.00         Matches:         7
Percent Similarity:       64.29%        Conservative:    2
Best Local Similarity:    50.00%        Mismatches:      5
Query Match:              48.48%        Indels:          0
DB:                       4             Gaps:            0

US-10-799-747-116 (1-20) x US-09-414-006-18 (1-1601)

Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||       |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802

RESULT 15
US-09-447-223-17/c
; Sequence 17, Application US/09447223
; Patent No. 6432664
;   GENERAL INFORMATION:
;        APPLICANT: Takashi OKADO et al.
;        TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
;                            AUREOBASIDIN SENSITIVITY
;        NUMBER OF SEQUENCES: 22
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Wenderoth, Lind & Ponack
;             STREET: 805 Fifteenth Street, N.W., #700
;             CITY: Washington
;             STATE: D.C.
```

; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,223
; FILING DATE: 23-No. 6432664-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: <Unknown>
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1601
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ANTI-SENSE: Yes
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-447-223-17

Alignment Scores:

Pred. No.:	68.2	Length:	1601
Score:	48.00	Matches:	7
Percent Similarity:	64.29%	Conservative:	2
Best Local Similarity:	50.00%	Mismatches:	5
Query Match:	48.48%	Indels:	0
DB:	4	Gaps:	0

US-10-799-747-116 (1-20) x US-09-447-223-17 (1-1601)

```
Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ||||||:::|||       |||   |||   ||||||   :::
Db        843 CATAGCATCCTTGATATTTTGGCCTGGGTTCCATATGGAGTC 802
```

Search completed: July 21, 2004, 02:54:35
Job time : 75 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 02:14:01 ; Search time 355 Seconds
(without alignments)
274.852 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 99
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3191023 seqs, 2439312756 residues

Total number of hits satisfying chosen parameters: 6382046

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161420_21560/app_query.fasta_1.199
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10799747_@CGN_1_1_723_@runat_19072004_161420_21560
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	98	99.0	1434	15	US-10-195-730-12	Sequence 12, Appl
2	57	57.6	87731	13	US-10-087-192-1342	Sequence 1342, Ap
3	56	56.6	8417	16	US-10-074-024-635	Sequence 635, App
4	53	53.5	5760	16	US-10-074-024-426	Sequence 426, App
5	53	53.5	5763	16	US-10-074-024-425	Sequence 425, App
c 6	51	51.5	2500	15	US-10-364-360-20	Sequence 20, Appl
c 7	51	51.5	9641	15	US-10-364-360-22	Sequence 22, Appl
c 8	51	51.5	56577	13	US-10-087-192-1396	Sequence 1396, Ap
c 9	51	51.5	2940917	13	US-10-027-632-174763	Sequence 174763,
c 10	51	51.5	2940917	16	US-10-027-632-174763	Sequence 174763,
c 11	50	50.5	620	12	US-10-152-319A-215	Sequence 215, App
c 12	50	50.5	740	13	US-10-027-632-11891	Sequence 11891, A
c 13	50	50.5	740	16	US-10-027-632-11891	Sequence 11891, A
14	50	50.5	1206	13	US-10-027-632-250731	Sequence 250731,
15	50	50.5	1206	13	US-10-027-632-250732	Sequence 250732,
16	50	50.5	1206	16	US-10-027-632-250731	Sequence 250731,
17	50	50.5	1206	16	US-10-027-632-250732	Sequence 250732,
c 18	50	50.5	3024	17	US-10-437-963-7778	Sequence 7778, Ap
19	49	49.5	519	13	US-10-027-632-284792	Sequence 284792,
20	49	49.5	519	16	US-10-027-632-284792	Sequence 284792,
21	49	49.5	793	13	US-10-027-632-144194	Sequence 144194,
22	49	49.5	793	13	US-10-027-632-144195	Sequence 144195,
23	49	49.5	793	16	US-10-027-632-144194	Sequence 144194,
24	49	49.5	793	16	US-10-027-632-144195	Sequence 144195,
c 25	49	49.5	808	13	US-10-027-632-155999	Sequence 155999,
c 26	49	49.5	808	16	US-10-027-632-155999	Sequence 155999,
27	49	49.5	828	13	US-10-027-632-158555	Sequence 158555,
28	49	49.5	828	13	US-10-027-632-158556	Sequence 158556,
29	49	49.5	828	16	US-10-027-632-158555	Sequence 158555,
30	49	49.5	828	16	US-10-027-632-158556	Sequence 158556,
31	49	49.5	1092	17	US-10-437-963-61207	Sequence 61207, A
c 32	49	49.5	2393	13	US-10-027-632-102504	Sequence 102504,
c 33	49	49.5	2393	16	US-10-027-632-102504	Sequence 102504,
c 34	49	49.5	32768	9	US-09-070-927A-128	Sequence 128, App
c 35	48.5	49.0	72705	13	US-10-087-192-1966	Sequence 1966, Ap
c 36	48	48.5	290	14	US-10-040-739-637	Sequence 637, App
37	48	48.5	375	10	US-09-803-719-2326	Sequence 2326, Ap
38	48	48.5	495	12	US-09-969-034-1082	Sequence 1082, Ap
39	48	48.5	587	13	US-10-424-599-90540	Sequence 90540, A
c 40	48	48.5	612	14	US-10-002-344A-83	Sequence 83, Appl
41	48	48.5	802	13	US-10-424-599-111198	Sequence 111198,
42	48	48.5	940	15	US-10-017-161-1223	Sequence 1223, Ap
43	48	48.5	940	16	US-10-292-798-1027	Sequence 1027, Ap
44	48	48.5	960	9	US-09-886-055-148	Sequence 148, App
45	48	48.5	960	10	US-09-804-291-148	Sequence 148, App

ALIGNMENTS

RESULT 1
US-10-195-730-12
; Sequence 12, Application US/10195730
; Publication No. US20030144492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: PZ017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-195-730-12

```
Alignment Scores:
Pred. No.:                6.04e-07        Length:          1434
Score:                    98.00           Matches:         19
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              98.99%          Indels:          0
DB:                       15              Gaps:            0

US-10-799-747-116 (1-20) x US-10-195-730-12 (1-1434)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563

RESULT 2
US-10-087-192-1342
; Sequence 1342, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1342
;   LENGTH: 87731
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-087-192-1342

Alignment Scores:
Pred. No.:                784             Length:          87731
Score:                    57.00           Matches:         10
Percent Similarity:       81.25%          Conservative:    3
Best Local Similarity:    62.50%          Mismatches:      3
Query Match:              57.58%          Indels:          0
DB:                       13              Gaps:            0

US-10-799-747-116 (1-20) x US-10-087-192-1342 (1-87731)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              ::::::||||||   |||   ||||||   ||||||||||:::|||
Db      41423 TCTAGTCATTCCCATCTTTGCTTTCTGAGCTGGACTCCTTACAGTCTG 41470

RESULT 3
US-10-074-024-635
; Sequence 635, Application US/10074024
; Publication No. US20030232975A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC001C1
;  CURRENT APPLICATION NUMBER: US/10/074,024
;  CURRENT FILING DATE:  2002-02-14
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 879
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 635
;   LENGTH: 8417
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2138)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (2140)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-074-024-635

Alignment Scores:
Pred. No.:                75.7            Length:          8417
Score:                    56.00           Matches:         9
Percent Similarity:       78.57%          Conservative:    2
Best Local Similarity:    64.29%          Mismatches:      3
Query Match:              56.57%          Indels:          0
DB:                       16              Gaps:            0

US-10-799-747-116 (1-20) x US-10-074-024-635 (1-8417)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrPro 14
              |||     ||||||:::|||:::||||||   |||||||||
Db        934 ATGCCACAACATTCGATTTTGACGTTTTTATTCTGGACCCCA 975

RESULT 4
US-10-074-024-426
; Sequence 426, Application US/10074024
; Publication No. US20030232975A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC001C1
;  CURRENT APPLICATION NUMBER: US/10/074,024
;  CURRENT FILING DATE:  2002-02-14
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 879
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 426
;   LENGTH: 5760
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-074-024-426

Alignment Scores:
Pred. No.:                159             Length:          5760
Score:                    53.00           Matches:         10
Percent Similarity:       68.75%          Conservative:    1
Best Local Similarity:    62.50%          Mismatches:      5
Query Match:              53.54%          Indels:          0
DB:                       16              Gaps:            0

US-10-799-747-116 (1-20) x US-10-074-024-426 (1-5760)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||   ||||||   |||||||||   ||||||      |||:::|||
Db        567 GCCCGGCACTCAACCCTGAGCTTTTGCCTTTGGGTTGTATATTCTCTT 614

RESULT 5
US-10-074-024-425
; Sequence 425, Application US/10074024
; Publication No. US20030232975A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC001C1
;  CURRENT APPLICATION NUMBER: US/10/074,024
;  CURRENT FILING DATE:  2002-02-14
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 879
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 425
;   LENGTH: 5763
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1871)
```

```
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-074-024-425

Alignment Scores:
Pred. No.:                159             Length:          5763
Score:                    53.00           Matches:         10
Percent Similarity:       68.75%          Conservative:    1
Best Local Similarity:    62.50%          Mismatches:      5
Query Match:              53.54%          Indels:          0
DB:                       16              Gaps:            0

US-10-799-747-116 (1-20) x US-10-074-024-425 (1-5763)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||   ||||||   |||||||||   ||||||      |||:::|||
Db        567 GCCCGGCACTCAACCCTGAGCTTTTGCCTTTGGGTTGTATATTCTCTT 614

RESULT 6
US-10-364-360-20/c
; Sequence 20, Application US/10364360
; Publication No. US20030180324A1
; GENERAL INFORMATION:
;  APPLICANT: GUERTLER, Lutz Gerhard
;  APPLICANT:  HAUSER, Hans Peter
;  APPLICANT:  DONGMO DELOKO, Yvette Beatrice
;  APPLICANT:  ZEKENG, Leopold
;  APPLICANT:  KAPTUE, Lazare
;  TITLE OF INVENTION: LENTIVIRUS FROM THE GROUP OF IMMUNODEFICIENCY VIRUSES OF DRILL MO
;  TITLE OF INVENTION:  (MANDRILLUS LEUCOPHAEUS) AND THEIR USE
;  FILE REFERENCE: 067595/0106
;  CURRENT APPLICATION NUMBER: US/10/364,360
;  CURRENT FILING DATE:  2003-02-12
;  PRIOR APPLICATION NUMBER: US/09/625,972
;  PRIOR FILING DATE: 2000-07-29
;  PRIOR APPLICATION NUMBER: DE 199 36 003.0
;  PRIOR FILING DATE: 1999-08-03
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
;   LENGTH: 2500
;   TYPE: DNA
;   ORGANISM: SIV - viral
US-10-364-360-20

Alignment Scores:
Pred. No.:                133             Length:          2500
Score:                    51.00           Matches:         8
Percent Similarity:       100.00%         Conservative:    3
Best Local Similarity:    72.73%          Mismatches:      0
Query Match:              51.52%          Indels:          0
DB:                       15              Gaps:            0

US-10-799-747-116 (1-20) x US-10-364-360-20 (1-2500)

Qy          7 LeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||||||||:::||||||:::||||||:::|||
Db       1453 TTGAGCTTTATACTTTGGTCCCCCTACTCCTTG 1421

RESULT 7
US-10-364-360-22/c
; Sequence 22, Application US/10364360
; Publication No. US20030180324A1
; GENERAL INFORMATION:
;  APPLICANT: GUERTLER, Lutz Gerhard
;  APPLICANT:  HAUSER, Hans Peter
;  APPLICANT:  DONGMO DELOKO, Yvette Beatrice
;  APPLICANT:  ZEKENG, Leopold
;  APPLICANT:  KAPTUE, Lazare
;  TITLE OF INVENTION: LENTIVIRUS FROM THE GROUP OF IMMUNODEFICIENCY VIRUSES OF DRILL MO
;  TITLE OF INVENTION:  (MANDRILLUS LEUCOPHAEUS) AND THEIR USE
;  FILE REFERENCE: 067595/0106
;  CURRENT APPLICATION NUMBER: US/10/364,360
;  CURRENT FILING DATE:  2003-02-12
;  PRIOR APPLICATION NUMBER: US/09/625,972
;  PRIOR FILING DATE: 2000-07-29
;  PRIOR APPLICATION NUMBER: DE 199 36 003.0
;  PRIOR FILING DATE: 1999-08-03
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 9641
;   TYPE: DNA
;   ORGANISM: SIV - viral
US-10-364-360-22

Alignment Scores:
Pred. No.:                639             Length:          9641
Score:                    51.00           Matches:         8
Percent Similarity:       100.00%         Conservative:    3
Best Local Similarity:    72.73%          Mismatches:      0
Query Match:              51.52%          Indels:          0
DB:                       15              Gaps:            0

US-10-799-747-116 (1-20) x US-10-364-360-22 (1-9641)

Qy          7 LeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||||||||:::||||||:::||||||:::|||
Db       1453 TTGAGCTTTATACTTTGGTCCCCCTACTCCTTG 1421

RESULT 8
US-10-087-192-1396/c
; Sequence 1396, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1396
;   LENGTH: 56577
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(56577)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1396

Alignment Scores:
Pred. No.:                5.03e+03        Length:          56577
Score:                    51.00           Matches:         10
Percent Similarity:       61.11%          Conservative:    1
Best Local Similarity:    55.56%          Mismatches:      7
Query Match:              51.52%          Indels:          0
DB:                       13              Gaps:            0

US-10-799-747-116 (1-20) x US-10-087-192-1396 (1-56577)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLys 18
              |||   ||||||      |||      ||||||||||||||||      |||:::
Db      29892 ATGATAGCTCACTGCAGTCTTGAACTCCTGCTTTGGACTCCCAAAGTGCTGGAA 29839

RESULT 9
US-10-027-632-174763/c
; Sequence 174763, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 174763
; LENGTH: 2940917
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2940917)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

Alignment Scores:

Pred. No.:	4.83e+05	Length:	2940917
Score:	51.00	Matches:	7
Percent Similarity:	84.62%	Conservative:	4
Best Local Similarity:	53.85%	Mismatches:	2
Query Match:	51.52%	Indels:	0
DB:	13	Gaps:	0

US-10-799-747-116 (1-20) x US-10-027-632-174763 (1-2940917)

```
Qy          3 AlaHisSerValLeuSerPheLeuLeuTrpThrProTyr 15
              :::||||||:::|||   |||:::   |||||||||:::
Db     189168 AGTCACTCCATTTTGTTTTTTGTTTTTTGGACTCCATTT 189130
```

RESULT 10
US-10-027-632-174763/c
; Sequence 174763, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 174763
; LENGTH: 2940917
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2940917)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

Alignment Scores:

Pred. No.:	4.83e+05	Length:	2940917
Score:	51.00	Matches:	7
Percent Similarity:	84.62%	Conservative:	4
Best Local Similarity:	53.85%	Mismatches:	2
Query Match:	51.52%	Indels:	0
DB:	16	Gaps:	0

US-10-799-747-116 (1-20) x US-10-027-632-174763 (1-2940917)

```
Qy          3 AlaHisSerValLeuSerPheLeuLeuTrpThrProTyr 15
              :::||||||:::|||   |||:::   |||||||||:::
Db     189168 AGTCACTCCATTTTGTTTTTTGTTTTTTGGACTCCATTT 189130
```

RESULT 11
US-10-152-319A-215/c
; Sequence 215, Application US/10152319A
; Publication No. US20040072160A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/10/152,319A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 215
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. AA891812
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(620)
; OTHER INFORMATION: n = a or c or g or t
US-10-152-319A-215

```
Alignment Scores:
Pred. No.:                38.7          Length:          620
Score:                    50.00         Matches:         10
Percent Similarity:       73.68%        Conservative:    4
Best Local Similarity:    52.63%        Mismatches:      5
Query Match:              50.51%        Indels:          0
DB:                       12            Gaps:            0

US-10-799-747-116 (1-20) x US-10-152-319A-215 (1-620)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              :::::::::|||   |||||||||   |||:::|||||||||      |||   |||
Db        234 CTGAGCTCTCACCTGGTCCTGAGTCCCCTCATCTGGACTCCTCCTCGTCTTCCTTCT 178

RESULT 12
US-10-027-632-11891/c
; Sequence 11891, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11891
;   LENGTH: 740
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-11891

Alignment Scores:
Pred. No.:                47.6          Length:          740
Score:                    50.00         Matches:         10
Percent Similarity:       84.62%        Conservative:    1
Best Local Similarity:    76.92%        Mismatches:      2
Query Match:              50.51%        Indels:          0
DB:                       13            Gaps:            0

US-10-799-747-116 (1-20) x US-10-027-632-11891 (1-740)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrPro 14
              |||:::   |||||||||||||||||||||||||   |||
Db        275 GCCTCYGTGAGTGTCTTAAGCTTCCTTCTCTGGGTTCCA 237

RESULT 13
US-10-027-632-11891/c
; Sequence 11891, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11891
;   LENGTH: 740
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-11891

Alignment Scores:
Pred. No.:                47.6          Length:          740
Score:                    50.00         Matches:         10
Percent Similarity:       84.62%        Conservative:    1
Best Local Similarity:    76.92%        Mismatches:      2
Query Match:              50.51%        Indels:          0
DB:                       16            Gaps:            0

US-10-799-747-116 (1-20) x US-10-027-632-11891 (1-740)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrPro 14
              |||:::   |||||||||||||||||||||||||   |||
Db        275 GCCTCYGTGAGTGTCTTAAGCTTCCTTCTCTGGGTTCCA 237

RESULT 14
US-10-027-632-250731
; Sequence 250731, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 250731
;   LENGTH: 1206
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-250731

Alignment Scores:
Pred. No.:                84.1          Length:          1206
Score:                    50.00         Matches:         8
Percent Similarity:       71.43%        Conservative:    2
```

```
Best Local Similarity:  57.14%          Mismatches:     4
Query Match:            50.51%          Indels:         0
DB:                     13              Gaps:           0

US-10-799-747-116 (1-20) x US-10-027-632-250731 (1-1206)

Qy          6 ValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              :::|||||||||||||   |||   ||||||   :::   |||
Db        704 ATTTTGAGCTTCCTATTCTGGCTACCATACCCCATCTCCAGC 745

RESULT 15
US-10-027-632-250732
; Sequence 250732, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 250732
;   LENGTH: 1206
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-250732

Alignment Scores:
Pred. No.:              84.1            Length:         1206
Score:                  50.00           Matches:        8
Percent Similarity:     71.43%          Conservative:   2
Best Local Similarity:  57.14%          Mismatches:     4
Query Match:            50.51%          Indels:         0
DB:                     13              Gaps:           0

US-10-799-747-116 (1-20) x US-10-027-632-250732 (1-1206)

Qy          6 ValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              :::|||||||||||||   |||   ||||||   :::   |||
Db        704 ATTTTGAGCTTCCTATTCTGGCTACCATACCCCATCTCCAGC 745


Search completed: July 21, 2004, 04:13:52
Job time : 1132 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 01:17:25 ; Search time 2469 Seconds
(without alignments)
241.897 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 99
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161416_21448/app_query.fasta_1.199
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747_@CGN_1_1_5180_@runat_19072004_161416_21448 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	98	99.0	240	10	BF910533	BF910533 CM4-UT004
2	98	99.0	575	12	BM722991	BM722991 UI-E-EO1-
3	98	99.0	765	12	BI914473	BI914473 603182264
4	98	99.0	785	12	BG484396	BG484396 602505037
5	98	99.0	1201	13	BX441923	BX441923 BX441923
6	56	56.6	352	9	AA654972	AA654972 nt61e09.s
c 7	55	55.6	642	29	CE789116	CE789116 tigr-gss-
8	55	55.6	723	28	AZ831423	AZ831423 2M0111H12
c 9	54	54.5	550	29	TA226G11P	AL480086 T. brucei
10	54	54.5	724	28	AQ645123	AQ645123 RPCI93-Ec
11	54	54.5	816	14	CA967853	CA967853 CcLX04a13
c 12	53	53.5	394	9	AA766276	AA766276 oa29c04.s
c 13	53	53.5	429	9	AA804444	AA804444 nz38c04.s
c 14	53	53.5	600	12	BI140807	BI140807 IP1_39_B0
15	53	53.5	603	12	BM571235	BM571235 fw75h06.x
16	53	53.5	635	13	BU262936	BU262936 603818292
17	53	53.5	641	13	BU350638	BU350638 603526071
18	53	53.5	641	13	BU448459	BU448459 603211744
19	53	53.5	642	13	BU353991	BU353991 603529626
20	53	53.5	643	13	BU394939	BU394939 603511454
21	53	53.5	649	13	BU337269	BU337269 603513967
22	53	53.5	657	13	BU390405	BU390405 603512306
23	53	53.5	784	29	CG759218	CG759218 ZMMBBb031
24	53	53.5	804	28	AZ906788	AZ906788 RPCI-24-1
25	53	53.5	904	29	CG768437	CG768437 TcB42.3_F
c 26	52.5	53.0	348	28	CC390782	CC390782 PUEHN29TD
27	52	52.5	254	9	AA324608	AA324608 EST27458
28	52	52.5	259	28	AQ104627	AQ104627 HS_3047_B
c 29	52	52.5	262	9	AI164098	AI164098 A054p15u
30	52	52.5	288	14	F07932	F07932 HSC2LH031 n
c 31	52	52.5	293	10	BB564042	BB564042 BB564042
c 32	52	52.5	372	29	CC717225	CC717225 OGLCD72TV
33	52	52.5	374	9	AA411283	AA411283 zv26f01.r
c 34	52	52.5	390	10	AW338878	AW338878 ha58e11.x
c 35	52	52.5	423	28	CC370277	CC370277 PUHGP23TD
c 36	52	52.5	584	9	AI754441	AI754441 cr25d04.x
c 37	52	52.5	607	9	AI672090	AI672090 ty63d04.x
38	52	52.5	653	28	CC004087	CC004087 PUDHO64TD
c 39	52	52.5	669	28	AZ572182	AZ572182 302PvC03
c 40	52	52.5	691	13	BU871645	BU871645 Q032H01 P
c 41	52	52.5	705	14	CD218841	CD218841 pgr1n.pk0
c 42	52	52.5	749	29	CG296747	CG296747 OG3CI17TV
c 43	52	52.5	759	12	BI837121	BI837121 603090085
c 44	52	52.5	759	28	BZ985367	BZ985367 PUECI80TD
45	52	52.5	837	29	CC717215	CC717215 OGLCD72TH

ALIGNMENTS

RESULT 1
BF910533/c
LOCUS BF910533 240 bp mRNA linear EST 18-JAN-2001
DEFINITION CM4-UT0042-011100-395-f08 UT0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF910533
VERSION BF910533.1 GI:12301991
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 240)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM4&t2=CM4-UT0042-011100-395-f08&t3=2000-11-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 81
High quality sequence stop: 240.

FEATURES Location/Qualifiers

source 1. .240
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="UT0042"
/note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Alignment Scores:

Pred. No.:	8.23e-06	Length:	240
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0
DB:	10	Gaps:	0

US-10-799-747-116 (1-20) x BF910533 (1-240)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 135
```

RESULT 2
BM722991
LOCUS BM722991 575 bp mRNA linear EST 01-MAR-2002
DEFINITION UI-E-EO1-aid-a-14-0-UI.r1 UI-E-EO1 Homo sapiens cDNA clone UI-E-EO1-aid-a-14-0-UI 5', mRNA sequence.
ACCESSION BM722991
VERSION BM722991.1 GI:19043963
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 575)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA sequence: 148-169, >AT_rich#Low_complexity (matched compliment)
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers

source 1. .575
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EO1-aid-a-14-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EO1"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EO1 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGCGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.:	2.54e-05	Length:	575
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0
DB:	12	Gaps:	0

US-10-799-747-116 (1-20) x BM722991 (1-575)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 510
```

RESULT 3
BI914473
LOCUS BI914473 765 bp mRNA linear EST 16-OCT-2001
DEFINITION 603182264F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246463 5', mRNA sequence.
ACCESSION BI914473
VERSION BI914473.1 GI:16178652
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 765)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

```
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11621  row: p  column: 16
            High quality sequence stop: 762.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5246463"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_121"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
                     fetal brains, female age 20 weeks, female age 24 weeks,
                     and male age 26 weeks.  Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     0.7-3.5 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 017. Note:
                     this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                3.67e-05        Length:          765
Score:                    98.00           Matches:         19
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              98.99%          Indels:          0
DB:                       12              Gaps:            0

US-10-799-747-116 (1-20) x BI914473 (1-765)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 612

RESULT 4
BG484396
LOCUS       BG484396                785 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602505037F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618473 5',
            mRNA sequence.
ACCESSION   BG484396
VERSION     BG484396.1  GI:13416675
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 785)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1376  row: j  column: 10
            High quality sequence start: 18
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1. .785
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4618473"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_77"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                3.79e-05        Length:          785
Score:                    98.00           Matches:         19
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              98.99%          Indels:          0
DB:                       12              Gaps:            0

US-10-799-747-116 (1-20) x BG484396 (1-785)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 378

RESULT 5
BX441923
LOCUS       BX441923                1201 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX441923 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF023YO02 5-PRIME, mRNA sequence.
ACCESSION   BX441923
VERSION     BX441923.1  GI:30771989
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2532.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF023BH01QP1&cluster=2532.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF023BH01QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF023YO02"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
```

/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:

Pred. No.:	6.56e-05	Length:	1201
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0
DB:	13	Gaps:	0

US-10-799-747-116 (1-20) x BX441923 (1-1201)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 202
```

RESULT 6
AA654972
LOCUS AA654972 352 bp mRNA linear EST 04-NOV-1997
DEFINITION nt61e09.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:1203016, mRNA sequence.
ACCESSION AA654972
VERSION AA654972.1 GI:2591126
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 352)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham.
FEATURES Location/Qualifiers
source 1. .352
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1203016"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr3"
/note="Vector: pAMP10; Site_1: Not1; Site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNAse-treated, total cellular RNA obtained from 5,000-10,000 microdissected cells histologically-determined to be fully malignant prostate cancer cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."
ORIGIN

Alignment Scores:

Pred. No.:	57.7	Length:	352
Score:	56.00	Matches:	9
Percent Similarity:	80.00%	Conservative:	3
Best Local Similarity:	60.00%	Mismatches:	3
Query Match:	56.57%	Indels:	0
DB:	9	Gaps:	0

US-10-799-747-116 (1-20) x AA654972 (1-352)

```
Qy          4 HisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLys 18
              ||||||   |||||||||:::::|||   |||   |||:::|||
Db         11 CACAGCAAATTATCATTCATTATTTGGGTGCCACAGGCCATTAAG 55
```

RESULT 7
CE789116/c
LOCUS CE789116 642 bp DNA linear GSS 30-SEP-2003
DEFINITION tigr-gss-dog-17000317786603 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE789116
VERSION CE789116.1 GI:37129879
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 642)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .642
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site_1: BstXI; Libraries were prepared from peripheral blood"
ORIGIN

Alignment Scores:

Pred. No.:	180	Length:	642
Score:	55.00	Matches:	10
Percent Similarity:	84.62%	Conservative:	1
Best Local Similarity:	76.92%	Mismatches:	2
Query Match:	55.56%	Indels:	0
DB:	29	Gaps:	0

US-10-799-747-116 (1-20) x CE789116 (1-642)

```
Qy          7 LeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||   |||:::   |||||||||
Db        208 TTATCATTCCTCCTTTGGGTTCCATTTTTATTAAAAAGC 170
```

RESULT 8
AZ831423
LOCUS AZ831423 723 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0111H12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0111H12 F, genomic survey sequence.
ACCESSION AZ831423

```
VERSION     AZ831423.1  GI:13001331
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0111  row: H  column: 12
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0111H12"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

Alignment Scores:
Pred. No.:                210             Length:         723
Score:                    55.00           Matches:        9
Percent Similarity:       66.67%          Conservative:   3
Best Local Similarity:    50.00%          Mismatches:     6
Query Match:              55.56%          Indels:         0
DB:                       28              Gaps:           0

US-10-799-747-116 (1-20) x AZ831423 (1-723)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLys 18
              |||   :::||||||   |||       |||:::|||   ||||||:::   |||
Db        121 ATGCTCTCTCACAGCACCTTGCAATTACTTGTGTGGTTACCCTATTCCACTAAG 174

RESULT 9
TA226G11P/c
LOCUS       TA226G11P               550 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 226g11, forward sequence,
            genomic survey sequence.
ACCESSION   AL480086
VERSION     AL480086.1  GI:11845965
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="226g11"
ORIGIN

Alignment Scores:
Pred. No.:                212             Length:         550
Score:                    54.00           Matches:        11
Percent Similarity:       73.33%          Conservative:   0
Best Local Similarity:    73.33%          Mismatches:     4
Query Match:              54.55%          Indels:         0
DB:                       29              Gaps:           0

US-10-799-747-116 (1-20) x TA226G11P (1-550)

Qy          3 AlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||       ||||||||||||||||   |||||||||   ||||||
Db        356 GCGGGTGGTGTGCTTAGTTTTTTGTTTTGGACTCCTTTAGCGCTG 312

RESULT 10
AQ645123
LOCUS       AQ645123                724 bp    DNA     linear   GSS 08-JUL-1999
DEFINITION  RPCI93-EcoRI-2M16.TV RPCI93-EcoRI Trypanosoma brucei genomic clone
            RPCI93-EcoRI-2M16, genomic survey sequence.
ACCESSION   AQ645123
VERSION     AQ645123.1  GI:5121833
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93
            Library for gene discovery and sequence-ready map construction
```

```
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI93-EcoRI-2M16.TP
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones and high density filters may be purchased from BACPAC
            Resources (http://bacpac.med.buffalo.edu. BAC end sequences search
            page: http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="RPCI93-EcoRI-2M16"
                     /clone_lib="RPCI93-EcoRI"
                     /note="Vector: pBACe3.6; Site_1: Eco RI; Site_2: Eco RI;
                     Constructed for The Institute for Genomic Research by
                     Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
                     Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
                     brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
                     partially digested with a combination of Eco RI and Eco RI
                     methylase (RPCI93-EcoRI segment) or Dpn II (RPCI93-DpnII
                     segment). High molecular weight fragments were ligated in
                     pBACe3.6 vector digested with Eco RI or Bam H1,
                     respectively. The average insert size is 141 Kb. Total
                     coverage (both segments): > 90 X the haploid
                     non-minichromosomal genome."
ORIGIN

Alignment Scores:
Pred. No.:                302             Length:         724
Score:                    54.00           Matches:        11
Percent Similarity:       73.33%          Conservative:   0
Best Local Similarity:    73.33%          Mismatches:     4
Query Match:              54.55%          Indels:         0
DB:                       28              Gaps:           0

US-10-799-747-116 (1-20) x AQ645123 (1-724)

Qy          3 AlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeu 17
              |||      |||||||||||||||   ||||||||   ||||||
Db        329 GCGGGTGGTGTGCTTAGTTTTTTGTTTTGGACTCCTTTAGCGCTG 373

RESULT 11
CA967853
LOCUS       CA967853                816 bp    mRNA    linear   EST 03-JAN-2003
DEFINITION  CcLX04a13k03f1 Carp mixed tissue library 1 Cyprinus carpio cDNA
            clone 13k03 5', mRNA sequence.
ACCESSION   CA967853
VERSION     CA967853.1  GI:27494410
KEYWORDS    EST.
SOURCE      Cyprinus carpio (common carp)
  ORGANISM  Cyprinus carpio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Cyprinus.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Gracey,A.Y., Fraser,E., Li,W. and Cossins,A.R.
  TITLE     Microarray and EST analysis of the carp (Cyprinus carpio)
            transcriptome during environmental stress
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Andrew R. Cossins
            Laboratory for Environmental Gene Regulation
            University of Liverpool
            School of Biological Sciences, The Biosciences Building, Crown
            Street, Liverpool, United Kingdom, L69 7ZB
            Tel: +44(0)151-795-4510
            Fax: +44(0)151-795-4431
            Email: cossins@liv.ac.uk
            Vector has been trimmed from this EST.
            Plate: 13  row: k  column: 03
            Seq primer: TriplEx 5' LD (5'-CTCGGGAAGCGCGCCATTGTGTTGGT-3')
            High quality sequence start: 33
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Cyprinus carpio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7962"
                     /clone="13k03"
                     /sex="Male & female"
                     /tissue_type="Skeletal white muscle, cardiac muscle,
                     kidney, brain, gill, intestinal mucosa"
                     /dev_stage="Adult"
                     /lab_host="E.coli Electromax DH10B"
                     /clone_lib="Carp mixed tissue library 1"
                     /note="Vector: pTriplEx2; Site_1: Sfi1 GGCCATTACGGCC;
                     Site_2: Sfi1 GGCCGCCTCGGCC; Normalized cDNA library
                     prepared from mixed tissues of warm, cold and hypoxia
                     challenged animals"
ORIGIN

Alignment Scores:
Pred. No.:                353             Length:         816
Score:                    54.00           Matches:        10
Percent Similarity:       80.00%          Conservative:   2
Best Local Similarity:    66.67%          Mismatches:     3
Query Match:              54.55%          Indels:         0
DB:                       14              Gaps:           0

US-10-799-747-116 (1-20) x CA967853 (1-816)

Qy          5 SerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              |||||||||   ||||||||||||   |||::::::|||   |||
Db        679 TCTGTCCTNCTTTTCTTACTTTGGGTNCCTTTCTCCCTTTCTTCC 723

RESULT 12
AA766276/c
LOCUS       AA766276                394 bp    mRNA    linear   EST 08-FEB-1998
DEFINITION  oa29c04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1306374 3'
            similar to contains element OFR repetitive element ;, mRNA
            sequence.
ACCESSION   AA766276
VERSION     AA766276.1  GI:2817514
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 666   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 357.
FEATURES             Location/Qualifiers
```

```
    source          1. .394
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1306374"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
                    /clone_lib="NCI_CGAP_GCB1"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                    primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                    ]. Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

Alignment Scores:
Pred. No.:                199             Length:         394
Score:                    53.00           Matches:        9
Percent Similarity:       80.00%          Conservative:   3
Best Local Similarity:    60.00%          Mismatches:     3
Query Match:              53.54%          Indels:         0
DB:                       9               Gaps:           0

US-10-799-747-116 (1-20) x AA766276 (1-394)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAla 16
              ||||||||||||   |||   |||:::||||||   |||::::::
Db        166 GCTGCACACTCTCCTCTTCTTTTTATTCTCTGGCCCCCATTTTCC 122

RESULT 13
AA804444/c
LOCUS       AA804444                429 bp    mRNA    linear   EST 18-FEB-1998
DEFINITION  nz38c04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1290054 3'
            similar to contains element OFR repetitive element ;, mRNA
            sequence.
ACCESSION   AA804444
VERSION     AA804444.1  GI:2875957
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 649   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 249.
FEATURES            Location/Qualifiers
     source         1. .429
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1290054"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
                    /clone_lib="NCI_CGAP_GCB1"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                    primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                    ]. Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN

Alignment Scores:
Pred. No.:                222             Length:         429
Score:                    53.00           Matches:        9
Percent Similarity:       80.00%          Conservative:   3
Best Local Similarity:    60.00%          Mismatches:     3
Query Match:              53.54%          Indels:         0
DB:                       9               Gaps:           0

US-10-799-747-116 (1-20) x AA804444 (1-429)

Qy          2 AlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAla 16
              ||||||||||||   |||   |||:::||||||   |||::::::
Db        166 GCTGCACACTCTCCTCTTCTTTTTATTCTCTGGCCCCCATTTTCC 122

RESULT 14
BI140807/c
LOCUS       BI140807                600 bp    mRNA    linear   EST 03-JUL-2001
DEFINITION  IP1_39_B03.b1_A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BI140807
VERSION     BI140807.1  GI:14593250
KEYWORDS    EST.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and
            Pratt,L.H.
  TITLE     An EST database from Sorghum: developing preanthesis pannicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for high quality sequence is
            20. Three-prime sequences, which are obtained with PolyTMix or T7
            sequencing primer, are presented as the reverse complement.
            Seq primer: JEN REV
            High quality sequence stop: 538
            POLYA=No.
FEATURES            Location/Qualifiers
     source         1. .600
                    /organism="Sorghum bicolor"
                    /mol_type="mRNA"
                    /cultivar="BTx623"
                    /db_xref="taxon:4558"
                    /clone_lib="Immature pannicle 1 (IP1)"
```

/note="Organ: Developing preanthesis pannicles; Vector: pBluescript II SK(-) from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

ORIGIN

Alignment Scores:

Pred. No.:	342	Length:	600
Score:	53.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	4
Best Local Similarity:	66.67%	Mismatches:	0
Query Match:	53.54%	Indels:	0
DB:	12	Gaps:	0

US-10-799-747-116 (1-20) x BI140807 (1-600)

```
Qy          8 SerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              :::||||||||||||||||||||:::|||::::::
Db        291 AATTTTCTGCTTTGGACTCCATATTCCTTAAGGACT 256
```

RESULT 15
BM571235
LOCUS BM571235 603 bp mRNA linear EST 21-FEB-2002
DEFINITION fw75h06.x1 Gong zebrafish testis Danio rerio cDNA clone IMAGE:5616251 3' similar to TR:O75130 O75130 KIAA0635 PROTEIN. ;, mRNA sequence.
ACCESSION BM571235
VERSION BM571235.1 GI:18853215
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 603)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
The library was constructed by Dr. Z. Gong. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. Please contact Zhiyuan Gong for further information on this library (National University of Singapore, Department of Biological Sciences, Lower Kent Ridge Road, Singapore 119260).
Seq primer: T7 from Gibco
High quality sequence stop: 329.

FEATURES Location/Qualifiers
source 1. .603
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5616251"
/sex="male"
/dev_stage="4-5 month"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Gong zebrafish testis"
/note="Organ: testis (pooled); Vector: pBluescript SK-; Site_1: XhoI; Site_2: EcoRI; Poly A+ RNA was isolatd from the testes of 31 male adult zebrafish (4-5 month old). cDNAs were made using oligo-dT primers and inserted into lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo mass-excised to pBluescript SK- following the Washington University protocol (http://genome.wustl.edu/est/lambda_protocol.shtml). Please contact Zhiyuan Gong for further information on this library (National University of Singapore, Department of Biological Sciences, Lower Kent Ridge Road, Singapore 119260)."

ORIGIN

Alignment Scores:

Pred. No.:	344	Length:	603
Score:	53.00	Matches:	11
Percent Similarity:	80.00%	Conservative:	1
Best Local Similarity:	73.33%	Mismatches:	3
Query Match:	53.54%	Indels:	0
DB:	12	Gaps:	0

US-10-799-747-116 (1-20) x BM571235 (1-603)

```
Qy          5 SerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||      ||||||||||||||||||    ||||||:::|||
Db        508 TCTGTAGCTCTTTTTCTCTTGTGGACACCAGCGGCTCTCCAGAGT 552
```

Search completed: July 21, 2004, 02:53:14
Job time : 2475 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 02:54:41 ; Search time 1704 Seconds
(without alignments)
508.721 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 20
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6933152

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161442_21981/app_query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUFFIX=olig.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747_@CGN_1_1_5600_@runat_19072004_161442_21981 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19	95.0	1434	6	BD078422	BD078422 101 human
2	19	95.0	3452	9	BC033650	BC033650 Homo sapi
3	19	95.0	3522	9	HSM804706	AL833393 Homo sapi
c 4	19	95.0	175081	9	AC021151	AC021151 Homo sapi
c 5	19	95.0	281662	2	AC117374	AC117374 Homo sapi
6	13	65.0	2020	9	HSM800227	AL049442 Homo sapi
c 7	9	45.0	153464	2	AC123991	AC123991 Mus muscu
c 8	9	45.0	167013	9	HSDJ828H9	AL121757 Human DNA
c 9	9	45.0	205725	10	AC122809	AC122809 Mus muscu
10	9	45.0	231768	2	AC110315	AC110315 Rattus no
c 11	9	45.0	254839	2	AC103176	AC103176 Rattus no
c 12	9	45.0	267352	2	AC098074	AC098074 Rattus no
c 13	9	45.0	306660	2	AC098226	AC098226 Rattus no
14	8	40.0	339	11	G25673	G25673 human STS E
15	8	40.0	356	11	G27758	G27758 human STS S
c 16	8	40.0	691	3	AK112944	AK112944 Ciona int
17	8	40.0	1503	6	AX654294	AX654294 Sequence
c 18	8	40.0	1524	9	AB070060	AB070060 Macaca fa
c 19	8	40.0	1751	9	AK130941	AK130941 Homo sapi
c 20	8	40.0	1791	8	AK106862	AK106862 Oryza sat
21	8	40.0	1806	10	MMY15799	Y15799 Mus musculu
22	8	40.0	1858	10	MMY15797	Y15797 Mus musculu
23	8	40.0	1916	10	MMY15800	Y15800 Mus musculu
24	8	40.0	1952	10	AF040749	AF040749 Mus muscu
c 25	8	40.0	1975	8	AY090039	AY090039 Nicotiana
26	8	40.0	1997	10	AF040747	AF040747 Mus muscu
27	8	40.0	1999	10	AF040748	AF040748 Mus muscu
28	8	40.0	2841	10	MMY15798	Y15798 Mus musculu
c 29	8	40.0	3200	8	AK119280	AK119280 Oryza sat
30	8	40.0	3628	8	AF238476	AF238476 Oryza sat
c 31	8	40.0	14782	9	HSHLADPB	X02228 Human HLA-D
c 32	8	40.0	37201	9	AP001215	AP001215 Homo sapi
c 33	8	40.0	52058	2	AC100335	AC100335 Mus muscu
c 34	8	40.0	56083	2	AC141162	AC141162 Rattus no
35	8	40.0	57630	2	AC100885	AC100885 Mus muscu
36	8	40.0	60504	2	AC090716	AC090716 Homo sapi
37	8	40.0	62263	2	AC100239	AC100239 Mus muscu
38	8	40.0	63139	10	AL833790	AL833790 Mouse DNA
c 39	8	40.0	64380	9	BX120009	BX120009 Human DNA
c 40	8	40.0	65209	2	AC101355	AC101355 Mus muscu
c 41	8	40.0	67358	2	AC108426	AC108426 Mus muscu
c 42	8	40.0	68790	9	AF017732	AF017732 Homo sapi
43	8	40.0	74658	8	OSA243961	AJ243961 Oryza sat
c 44	8	40.0	91585	10	AL606661	AL606661 Mouse DNA
c 45	8	40.0	93842	9	AL845446	AL845446 Human DNA

ALIGNMENTS

RESULT 1

```
BD078422
LOCUS       BD078422                1434 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  101 human secretory proteins.
ACCESSION   BD078422
VERSION     BD078422.1  GI:22624025
KEYWORDS    JP 2001519156-A/11.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1434)
  AUTHORS   Duan,R.D., Florence,K.A., Rosen,C.A., Ruben,S.M., Greene,J.M.,
            Young,P., Ferrie,A.M., Yu,G.L., Janat,F., Ni,J., Carter,K.C.,
            Endress,G.A., Feng,P., Lafleur,D.W. and Shi,Y.
  TITLE     101 human secretory proteins
  JOURNAL   Patent: JP 2001519156-A 11 23-OCT-2001;
            HUMAN GENOME SCIENCES INC
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001519156-A/11
            PD   23-OCT-2001
            PF   01-OCT-1998 JP 2000515006
            PR   02-OCT-1997 US    60/060837,02-OCT-1997 US    60/060862 PR
            02-OCT-1997 US    60/060839,02-OCT-1997 US    60/060866 PR
            02-OCT-1997 US    60/060843,02-OCT-1997 US    60/060836 PR
            02-OCT-1997 US    60/060838,02-OCT-1997 US    60/060874 PR
            02-OCT-1997 US    60/060833,02-OCT-1997 US    60/060884 PR
            02-OCT-1997 US    60/060880
            PI   ROXANNE D DUAN,KIMBERLY A FLORENCE,CRAIG A ROSEN,STEVEN M  PI
             RUBEN,
            PI   JOHN M GREENE,PAUL YOUNG,ANN M FERRIE,GUO
            LIANG YU,FOUAD JANAT,
            PI   JIAN NI,
            PI   KENNETH C CARTER,GREGORY A ENDRESS,PING FENG,DAVID W LAFLEUR,
            PI   YANGGU SHI
            PC   C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P25/00,A61P35/00,
            PC   A61P37/00,
            PC   A61P43/00,C07K14/47,C07K16/00,C12N5/10,C12P21/02,C12Q1/68, PC
             G01N33/53,
            PC   G01N33/53,G01N33/566,C12N15/00,A61K37/02,C12N5/00 CC   101
            human secretory proteins
            FH   Key             Location/Qualifiers
            FT   source          1. .1434
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .1434
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                3.07e-10        Length:          1434
Score:                    19.00           Matches:         19
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              95.00%          Indels:          0
DB:                       6               Gaps:            0

US-10-799-747-116 (1-20) x BD078422 (1-1434)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563

RESULT 2
BC033650
LOCUS       BC033650                3452 bp    mRNA    linear   PRI 06-OCT-2003
DEFINITION  Homo sapiens hypothetical protein FLJ14431, mRNA (cDNA clone
            MGC:44889 IMAGE:5574637), complete cds.
ACCESSION   BC033650
VERSION     BC033650.1  GI:21707066
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3452)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 3452)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 69 Row: n Column: 19
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 14249445.
FEATURES             Location/Qualifiers
     source          1. .3452
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:44889 IMAGE:5574637"
                     /tissue_type="Duodenum, adenocarcinoma"
                     /clone_lib="NIH_MGC_88"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .3452
                     /gene="FLJ14431"
                     /db_xref="LocusID:84869"
     CDS             159. .872
```

```
                     /codon_start=1
                     /product="FLJ14431 protein"
                     /protein_id="AAH33650.1"
                     /db_xref="GI:21707067"
                     /db_xref="LocusID:84869"
                     /translation="MDKVCAVFGGSRGIGRAVAQLMARKGYRLAVIARNLEGAKAAAG
                     DLGGDHLAFSCDVAKEHDVQNTFEELEKHLGRVNFLVNAAGINRDGLLVRTKTEDMVS
                     QLHTNLLGSMLTCKAAMRTMIQQQGGSIVNVGSIVGLKGNSGQSVYSASKGGLVGFSR
                     ALAKEVARKKIRVNVVAPGFVHTDMTKDLKEEHLKKNIPLGRFGETIEVAHAVVFLLE
                     SPYITGHVLVVDGGLQLIL"
     misc_feature    159. .857
                     /note="adh_short; Region: short chain dehydrogenase. This
                     family contains a wide variety of dehydrogenases"
                     /db_xref="CDD:pfam00106"
ORIGIN

Alignment Scores:
Pred. No.:              6.82e-10        Length:         3452
Score:                  19.00           Matches:        19
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            95.00%          Indels:         0
DB:                     9               Gaps:           0

US-10-799-747-116 (1-20) x BC033650 (1-3452)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2542 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 2598

RESULT 3
HSM804706
LOCUS       HSM804706               3522 bp    mRNA    linear   PRI 10-JUL-2002
DEFINITION  Homo sapiens mRNA; cDNA DKFZp762K109 (from clone DKFZp762K109).
ACCESSION   AL833393
VERSION     AL833393.1  GI:21734029
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3522)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp762K109) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .3522
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp762K109"
                     /db_xref="taxon:9606"
                     /clone="DKFZp762K109"
                     /tissue_type="melanoma (MeWo cell line)"
                     /clone_lib="762 (synonym: hmel2). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     polyA_signal    3420. .3425
     polyA_site      3440
ORIGIN

Alignment Scores:
Pred. No.:              6.94e-10        Length:         3522
Score:                  19.00           Matches:        19
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            95.00%          Indels:         0
DB:                     9               Gaps:           0

US-10-799-747-116 (1-20) x HSM804706 (1-3522)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2552 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 2608

RESULT 4
AC021151/c
LOCUS       AC021151               175081 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-483A20 from 4, complete sequence.
ACCESSION   AC021151
VERSION     AC021151.8  GI:15145598
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175081)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 175081)
  AUTHORS   Kyung,K. and Abbott,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-483A20
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 175081)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 175081)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 175081)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 9, 2001 this sequence version replaced gi:13877272.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0483A20
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
```

McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-36G9. Actual start of this clone is at base position 1 of RP11-483A20; actual end is at base position 175081 of RP11-483A20.

FEATURES	Location/Qualifiers
source	1. .175081
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-483A20"
	/clone_lib="RPCI-11"
repeat_region	1. .3267
	/rpt_family="L1"
repeat_region	3270. .3453
	/rpt_family="Alu"
repeat_region	3434. .3464
	/rpt_family="(A)n"
repeat_region	3455. .4161
	/rpt_family="L1"
misc_feature	7949. .7956
	/note="match to EST AW006057 (NID:g5854835) wz81d04.x1"
repeat_region	10197. .10489
	/rpt_family="L1"
repeat_region	10480. .10501
	/rpt_family="AT_rich"
repeat_region	10495. .10665
	/rpt_family="L1"
repeat_region	10731. .11051
	/rpt_family="L1"
repeat_region	11052. .11576
	/rpt_family="L1"
repeat_region	11564. .11598
	/rpt_family="(A)n"
repeat_region	11642. .12324
	/rpt_family="L1"
repeat_region	11769. .11794
	/rpt_family="AT_rich"
repeat_region	11939. .11962
	/rpt_family="AT_rich"
repeat_region	12335. .12412
	/rpt_family="L1"
repeat_region	12392. .12621
	/rpt_family="L1"
repeat_region	12788. .13096
	/rpt_family="Alu"
repeat_region	13069. .13102
	/rpt_family="(CAAAA)n"
repeat_region	13325. .13667
	/rpt_family="Alu"
repeat_region	13444. .13484
	/rpt_family="(TAAA)n"
repeat_region	13640. .13686
	/rpt_family="(GAAAA)n"
repeat_region	14249. .14422
	/rpt_family="MER1_type"
repeat_region	14980. .15182
	/rpt_family="L1"
repeat_region	15260. .15482
	/rpt_family="L1"
repeat_region	15483. .15985
	/rpt_family="MER1_type"
repeat_region	15986. .16105
	/rpt_family="L1"
repeat_region	16117. .16251
	/rpt_family="L1"
repeat_region	16288. .16356
	/rpt_family="L1"
repeat_region	16373. .16545
	/rpt_family="L1"
repeat_region	16550. .18773
	/rpt_family="L1"
repeat_region	18770. .20218
	/rpt_family="L1"
repeat_region	19228. .19346
	/rpt_family="T-rich"
repeat_region	20219. .20514
	/rpt_family="Alu"
repeat_region	20515. .22484
	/rpt_family="L1"
repeat_region	22584. .22882
	/rpt_family="Alu"
repeat_region	22748. .22768
	/rpt_family="AT_rich"
repeat_region	22883. .23460
	/rpt_family="L1"
repeat_region	23461. .23747
	/rpt_family="MaLR"
repeat_region	23749. .24514
	/rpt_family="L1"
misc_feature	24456
	/note="match to EST AW902102 (NID:g8066307)"
misc_feature	24488. .24797
	/note="match to EST AW902102 (NID:g8066307)"
repeat_region	24576. .24625
	/rpt_family="Mariner"
repeat_region	25484. .25787
	/rpt_family="Alu"
repeat_region	25758. .25787
	/rpt_family="(A)n"
repeat_region	25796. .26014
	/rpt_family="MIR"
repeat_region	26657. .26889
	/rpt_family="L1"
repeat_region	26890. .27205
	/rpt_family="MaLR"
repeat_region	27206. .27439
	/rpt_family="Alu"
repeat_region	27440. .27490
	/rpt_family="MaLR"
repeat_region	27491. .28717
	/rpt_family="L1"
repeat_region	28718. .28742
	/rpt_family="(T)n"
misc_feature	28735. .28771
	/note="match to EST AA954400 (NID:g3118095) oo01f01.s1"
misc_feature	28747. .28826
	/note="match to EST AW006057 (NID:g5854835) wz81d04.x1"
repeat_region	29022. .29308
	/rpt_family="L1"
misc_feature	29384. .30105
	/note="match to EST BF680971 (NID:g11954866)"
misc_feature	29687. .30351
	/note="match to EST BG429938 (NID:g13336444)"

Alignment Scores:

Pred. No.:	2.42e-08	Length:	175081
Score:	19.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	95.00%	Indels:	0

```
DB:                 9          Gaps:          0

US-10-799-747-116 (1-20) x AC021151 (1-175081)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57641 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 57585

RESULT 5
AC117374/c
LOCUS       AC117374              281662 bp    DNA     linear   HTG 02-MAY-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
            PROGRESS ***, 48 unordered pieces.
ACCESSION   AC117374 AC011694
VERSION     AC117374.1  GI:20127352
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 281662)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 281662)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 281662)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28, 2002 this sequence version replaced gi:10045385.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCND
                Center clone name: RP11-19D19
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 273063 bases at least Q40
                Consensus quality: 294336 bases at least Q30
                Consensus quality: 306942 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 48 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2464: contig of 2464 bp in length
            *     2465     2564: gap of unknown length
            *     2565     4603: contig of 2039 bp in length
            *     4604     4703: gap of unknown length
            *     4704     7113: contig of 2410 bp in length
            *     7114     7213: gap of unknown length
            *     7214     9517: contig of 2304 bp in length
            *     9518     9617: gap of unknown length
            *     9618    12107: contig of 2490 bp in length
            *    12108    12207: gap of unknown length
            *    12208    15014: contig of 2807 bp in length
            *    15015    15114: gap of unknown length
            *    15115    17831: contig of 2717 bp in length
            *    17832    17931: gap of unknown length
            *    17932    20480: contig of 2549 bp in length
            *    20481    20580: gap of unknown length
            *    20581    22700: contig of 2120 bp in length
            *    22701    22800: gap of unknown length
            *    22801    24909: contig of 2109 bp in length
            *    24910    25009: gap of unknown length
            *    25010    27313: contig of 2304 bp in length
            *    27314    27413: gap of unknown length
            *    27414    29798: contig of 2385 bp in length
            *    29799    29898: gap of unknown length
            *    29899    33466: contig of 3568 bp in length
            *    33467    33566: gap of unknown length
            *    33567    36679: contig of 3113 bp in length
            *    36680    36779: gap of unknown length
            *    36780    40472: contig of 3693 bp in length
            *    40473    40572: gap of unknown length
            *    40573    42683: contig of 2111 bp in length
            *    42684    42783: gap of unknown length
            *    42784    46319: contig of 3536 bp in length
            *    46320    46419: gap of unknown length
            *    46420    49507: contig of 3088 bp in length
            *    49508    49607: gap of unknown length
            *    49608    53212: contig of 3605 bp in length
            *    53213    53312: gap of unknown length
            *    53313    57267: contig of 3955 bp in length
            *    57268    57367: gap of unknown length
            *    57368    60444: contig of 3077 bp in length
            *    60445    60544: gap of unknown length
            *    60545    65859: contig of 5315 bp in length
            *    65860    65959: gap of unknown length
```

```
            *    65960    69038: contig of 3079 bp in length
            *    69039    69138: gap of unknown length
            *    69139    73445: contig of 4307 bp in length
            *    73446    73545: gap of unknown length
            *    73546    78107: contig of 4562 bp in length
            *    78108    78207: gap of unknown length
            *    78208    82504: contig of 4297 bp in length
            *    82505    82604: gap of unknown length
            *    82605    86976: contig of 4372 bp in length
            *    86977    87076: gap of unknown length
            *    87077    91315: contig of 4239 bp in length
            *    91316    91415: gap of unknown length
            *    91416    95084: contig of 3669 bp in length
            *    95085    95184: gap of unknown length
            *    95185   102348: contig of 7164 bp in length
            *   102349   102448: gap of unknown length
            *   102449   105973: contig of 3525 bp in length
            *   105974   106073: gap of unknown length
            *   106074   112891: contig of 6818 bp in length
            *   112892   112991: gap of unknown length
            *   112992   116221: contig of 3230 bp in length
            *   116222   116321: gap of unknown length
            *   116322   119798: contig of 3477 bp in length
            *   119799   119898: gap of unknown length
            *   119899   124870: contig of 4972 bp in length
            *   124871   124970: gap of unknown length
            *   124971   130790: contig of 5820 bp in length
            *   130791   130890: gap of unknown length
            *   130891   134639: contig of 3749 bp in length
            *   134640   134739: gap of unknown length
            *   134740   140637: contig of 5898 bp in length
            *   140638   140737: gap of unknown length
            *   140738   145541: contig of 4804 bp in length
            *   145542   145641: gap of unknown length
            *   145642   155057: contig of 9416 bp in length
            *   155058   155157: gap of unknown length
            *   155158   164357: contig of 9200 bp in length
            *   164358   164457: gap of unknown length
            *   164458   174892: contig of 10435 bp in length
            *   174893   174992: gap of unknown length
            *   174993   184400: contig of 9408 bp in length
            *   184401   184500: gap of unknown length
            *   184501   198385: contig of 13885 bp in length
            *   198386   198485: gap of unknown length
            *   198486   216436: contig of 17951 bp in length
            *   216437   216536: gap of unknown length
            *   216537   232916: contig of 16380 bp in length
            *   232917   233016: gap of unknown length
            *   233017   255437: contig of 22421 bp in length
            *   255438   255537: gap of unknown length
            *   255538   281662: contig of 26125 bp in length.
FEATURES             Location/Qualifiers
     source          1. .281662
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-19D19"
ORIGIN

Alignment Scores:
Pred. No.:              3.73e-08        Length:         281662
Score:                  19.00           Matches:        19
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            95.00%          Indels:         0
DB:                     2               Gaps:           0

US-10-799-747-116 (1-20) x AC117374 (1-281662)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145718 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCATATGCTCTGAAATCA 145662

RESULT 6
HSM800227
LOCUS       HSM800227              2020 bp    mRNA    linear   PRI 18-FEB-2000
DEFINITION  Homo sapiens mRNA; cDNA DKFZp586N1720 (from clone DKFZp586N1720).
ACCESSION   AL049442
VERSION     AL049442.1  GI:4500222
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2020)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp586N1720) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .2020
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp586N1720"
                     /db_xref="taxon:9606"
                     /clone="DKFZp586N1720"
                     /tissue_type="uterus"
                     /clone_lib="586 (synonym: hute1). Vector pSport1; host
                     DH10B; sites NotI + SalI/MluI"
                     /dev_stage="adult"
     polyA_signal    1983. .1988
     polyA_site      2003
ORIGIN

Alignment Scores:
Pred. No.:              0.000538        Length:         2020
Score:                  13.00           Matches:        13
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            65.00%          Indels:         0
DB:                     9               Gaps:           0

US-10-799-747-116 (1-20) x HSM800227 (1-2020)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThr 13
              |||||||||||||||||||||||||||||||||||||||
Db       1115 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACA 1153

RESULT 7
AC123991/c
LOCUS       AC123991               153464 bp    DNA     linear   HTG 17-MAR-2003
DEFINITION  Mus musculus clone RP24-137I10, WORKING DRAFT SEQUENCE, 3 unordered
            pieces.
ACCESSION   AC123991
VERSION     AC123991.3  GI:28975888
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 153464)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-137I10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153464)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
```

Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 153464)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Mar 17, 2003 this sequence version replaced gi:28201678.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L26544
Center clone name: 137_I_10
---------- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 152581 bases at least Q40
Consensus quality: 152929 bases at least Q30
Consensus quality: 153074 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 153264; sum-of-contigs
Quality coverage: 11.8 in Q20 bases; agarose-fp
Quality coverage: 11.7 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 132252: contig of 132252 bp in length
* 132253 132352: gap of 100 bp
* 132353 134109: contig of 1757 bp in length
* 134110 134209: gap of 100 bp
* 134210 153464: contig of 19255 bp in length.

FEATURES Location/Qualifiers

source	1. .153464
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/clone="RP24-137I10"
	/clone_lib="RPCI-24 Male Mouse BAC"
misc_feature	1. .132252
	/note="assembly_fragment
	clone_end:SP6
	vector_side:left"
misc_feature	132353. .134109
	/note="assembly_fragment"
misc_feature	134210. .153464
	/note="assembly_fragment
	clone_end:T7
	vector_side:right"

ORIGIN

Alignment Scores:

Pred. No.:	326	Length:	153464
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.00%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AC123991 (1-153464)

```
Qy          3 AlaHisSerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||||||||||
Db     102764 GCCCACTCAGTACTGAGCTTTTTATTG 102738
```

RESULT 8
HSDJ828H9/c
LOCUS HSDJ828H9 167013 bp DNA linear PRI 06-MAR-2001
DEFINITION Human DNA sequence from clone RP5-828H9 on chromosome 20. Contains two putative novel genes, STSs, GSSs and CpG islands, complete sequence.
ACCESSION AL121757
VERSION AL121757.7 GI:6114793
KEYWORDS HTG; CpG island.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167013)
AUTHORS Blakey,S.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Oct 25, 1999 this sequence version replaced gi:6102834.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

```
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP5-828H9 is from the
            library RPCI-5 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP5-828H9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP4-680N4 is at 166918 in this sequence.
            The true right end of clone RP5-1022P6 is at 99 in this sequence.
            The true right end of clone RP4-766D4 is at 46185 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .167013
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP5-828H9"
                     /clone_lib="RPCI-5"
     repeat_region   785. .1214
                     /note="HERV17 repeat: matches 6052. .6478 of consensus"
     misc_feature    1271. .1441
                     /note="match: GSS: Em:AQ780632"
     repeat_region   1336. .1566
                     /note="MER66-internal repeat: matches 4624. .4847 of
                     consensus"
     misc_feature    1446. .1607
                     /note="match: GSS: Em:AQ780632"
     repeat_region   1798. .1923
                     /note="MER51-internal repeat: matches 7690. .7816 of
                     consensus"
     repeat_region   2182. .2263
                     /note="MER51B repeat: matches 190. .523 of consensus"
     repeat_region   2375. .2451
                     /note="AluS repeat: matches 1. .67 of consensus"
     repeat_region   2455. .2627
                     /note="L1M4 repeat: matches 5635. .5810 of consensus"
     repeat_region   2628. .2935
                     /note="AluSq repeat: matches 1. .308 of consensus"
     repeat_region   2936. .3001
                     /note="L1M4 repeat: matches 5569. .5635 of consensus"
     repeat_region   3002. .3305
                     /note="AluJb repeat: matches 1. .300 of consensus"
     repeat_region   3306. .3764
                     /note="L1M4 repeat: matches 5120. .5569 of consensus"
     repeat_region   3765. .4077
                     /note="AluJo repeat: matches 1. .292 of consensus"
     repeat_region   4078. .4223
                     /note="L1M4 repeat: matches 4973. .5120 of consensus"
     repeat_region   4224. .4361
                     /note="AluJo/FRAM repeat: matches 158. .294 of consensus"
     gene            complement(join(4666. .4827,5364. .5428,7817. .7945))
                     /gene="dJ828H9.2"
     mRNA            complement(join(4666. .4827,5364. .5428,7817. .>7945))
                     /gene="dJ828H9.2"
                     /product="dJ828H9.2 (putative novel transcript)"
                     /note="match: ESTs: Em:AI218423"
                     /evidence=not_experimental
     polyA_site      complement(4666)
                     /gene="dJ828H9.2"
     repeat_region   5065. .5364
                     /note="AluSx repeat: matches 1. .298 of consensus"
     repeat_region   5519. .5794
                     /note="L2 repeat: matches 2405. .2680 of consensus"
     misc_feature    5934. .6244
                     /note="match: GSS: Em:AQ415067"
     misc_feature    6021. .6321
                     /note="match: GSS: Em:AQ531024"
     repeat_region   6258. .7406
                     /note="L1PA5 repeat: matches 4995. .6144 of consensus"
     repeat_region   8230. .8343
                     /note="MIR repeat: matches 45. .166 of consensus"
     gene            complement(join(9193. .9766,11907. .12175))
                     /gene="dJ828H9.1"
     mRNA            complement(join(9193. .9766,11907. .>12175))
                     /gene="dJ828H9.1"
                     /product="dJ828H9.1 (putative novel transcript)"
                     /note="match: ESTs: Em:AA431369 Em:AA382348 Em:AA397834
                     Em:AA758864 Em:AA906363 Em:AA757279 Em:AW341461
                     Em:AA432366"
                     /evidence=not_experimental
     polyA_site      complement(9193)
                     /gene="dJ828H9.1"
     polyA_site      complement(9426)
                     /gene="dJ828H9.1"
     repeat_region   10546. .10649
                     /note="L2 repeat: matches 541. .654 of consensus"
     repeat_region   10708. .10784
                     /note="ORSL repeat: matches 206. .275 of consensus"
     repeat_region   10785. .11076
                     /note="AluY repeat: matches 2. .295 of consensus"
     repeat_region   11077. .11230
                     /note="ORSL repeat: matches 275. .458 of consensus"
     misc_feature    complement(12059. .12555)
                     /note="match: GSS: Em:AQ276946"
     repeat_region   13524. .13668
                     /note="MER5B repeat: matches 17. .166 of consensus"
     repeat_region   14312. .14516
                     /note="AluJb repeat: matches 111. .306 of consensus"
     repeat_region   14691. .15441
                     /note="L1MB3 repeat: matches 5264. .6016 of consensus"
     repeat_region   15414. .15494
                     /note="L1ME3A repeat: matches 6025. .6104 of consensus"
     repeat_region   15495. .15805
                     /note="AluSq repeat: matches 1. .309 of consensus"
     repeat_region   15976. .16160
                     /note="MLT1G repeat: matches 39. .229 of consensus"
     repeat_region   16253. .16343
                     /note="MER5A repeat: matches 58. .145 of consensus"
     repeat_region   16275. .16389
                     /note="MER5A repeat: matches 1. .123 of consensus"
     misc_feature    16406. .16912
                     /note="match: GSS: Em:AQ792876"
     repeat_region   16434. .16597
                     /note="MIR repeat: matches 91. .262 of consensus"
     misc_feature    17457. .17911
                     /note="match: GSS: Em:AQ135214"
     misc_feature    17487. .17994
                     /note="match: GSS: Em:AQ215088"
     repeat_region   17672. .18210
                     /note="MER52A repeat: matches 1207. .1755 of consensus"
     repeat_region   18311. .18921
                     /note="MER52A repeat: matches 660. .1349 of consensus"
     repeat_region   18976. .19568
                     /note="MER4-internal repeat: matches 4586. .5102 of
                     consensus"
     repeat_region   19578. .24868
                     /note="L1PA8 repeat: matches 10. .5050 of consensus"
     repeat_region   24869. .25168
                     /note="AluSq repeat: matches 1. .300 of consensus"
     repeat_region   25169. .26305
                     /note="L1PA8 repeat: matches 5050. .6163 of consensus"
     repeat_region   26308. .26385
```

```
                     /note="MER4-internal repeat: matches 3313. .3393 of
                     consensus"
     repeat_region   26386. .26686
                     /note="AluJb repeat: matches 1. .302 of consensus"
     repeat_region   26687. .26832
                     /note="MER4-internal repeat: matches 3393. .3543 of
                     consensus"
     repeat_region   26833. .27129
                     /note="AluSx repeat: matches 1. .296 of consensus"
     repeat_region   27130. .27559
                     /note="MER4-internal repeat: matches 3543. .3989 of
                     consensus"
     repeat_region   27560. .27865
                     /note="AluJb repeat: matches 1. .301 of consensus"
     repeat_region   27866. .28012
                     /note="MER4-internal repeat: matches 3989. .4140 of
                     consensus"
     repeat_region   28013. .28262
                     /note="AluJo repeat: matches 3. .264 of consensus"
     repeat_region   28263. .28514
                     /note="MER4-internal repeat: matches 4140. .4401 of
                     consensus"
     repeat_region   28515. .28797
                     /note="AluSx repeat: matches 1. .284 of consensus"
     repeat_region   28798. .29064
                     /note="MER4-internal repeat: matches 4401. .6347 of
                     consensus"
     misc_feature    28993. .29239

Alignment Scores:
Pred. No.:                353           Length:         167013
Score:                    9.00          Matches:        9
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              45.00%        Indels:         0
DB:                       9             Gaps:           0

US-10-799-747-116 (1-20) x HSDJ828H9 (1-167013)

Qy          5 SerValLeuSerPheLeuLeuTrpThr 13
              |||||||||||||||||||||||||||
Db      84047 TCTGTGCTCAGCTTTCTTCTCTGGACA 84021

RESULT 9
AC122809/c
LOCUS       AC122809               205725 bp    DNA     linear   ROD 25-NOV-2003
DEFINITION  Mus musculus BAC clone RP23-272E5 from chromosome 7, complete
            sequence.
ACCESSION   AC122809
VERSION     AC122809.4  GI:33300789
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 205725)
  AUTHORS   Swearengen-Shahid,S. and Meyer,R.
  TITLE     The sequence of Mus musculus BAC clone RP23-272E5
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 205725)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 205725)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 205725)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 205725)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 205725)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 29, 2003 this sequence version replaced gi:28372763.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0272E05
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES                Location/Qualifiers
     source             1. .205725
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="7"
                        /map="7"
                        /clone="RP23-272E5"
                        /clone_lib="RPCI-23"
     repeat_region      1. .94
                        /rpt_family="Alu"
     repeat_region      92. .106
                        /rpt_family="MaLR"
     repeat_region      107. .295
                        /rpt_family="B2"
     repeat_region      296. .525
                        /rpt_family="MaLR"
     repeat_region      706. .778
                        /rpt_family="Alu"
     repeat_region      831. .960
                        /rpt_family="MIR"
     repeat_region      1158. .1200
                        /rpt_family="Alu"
     repeat_region      1201. .1596
```

```
                     /rpt_family="MaLR"
     repeat_region   1597. .1667
                     /rpt_family="Alu"
     repeat_region   1817. .1930
                     /rpt_family="Alu"
     repeat_region   2015. .2069
                     /rpt_family="ID"
     repeat_region   2694. .2809
                     /rpt_family="Alu"
     repeat_region   2866. .3093
                     /rpt_family="B2"
     repeat_region   3155. .3231
                     /rpt_family="ERV1"
     repeat_region   3302. .3431
                     /rpt_family="Alu"
     repeat_region   4114. .4317
                     /rpt_family="L1"
     repeat_region   4320. .4395
                     /rpt_family="L1"
     repeat_region   4396. .4542
                     /rpt_family="Alu"
     repeat_region   4543. .4609
                     /rpt_family="L1"
     repeat_region   4610. .4759
                     /rpt_family="Alu"
     repeat_region   4760. .4898
                     /rpt_family="L1"
     repeat_region   4899. .5280
                     /rpt_family="MaLR"
     repeat_region   5281. .6123
                     /rpt_family="L1"
     repeat_region   6129. .6149
                     /rpt_family="MaLR"
     repeat_region   6150. .6253
                     /rpt_family="Alu"
     repeat_region   6340. .6478
                     /rpt_family="Alu"
     repeat_region   8595. .8732
                     /rpt_family="Alu"
     repeat_region   8626. .8823
                     /rpt_family="MaLR"
     repeat_region   9217. .9562
                     /rpt_family="MaLR"
     repeat_region   15551. .15699
                     /rpt_family="Alu"
     repeat_region   15734. .15835
                     /rpt_family="Alu"
     repeat_region   16956. .17076
                     /rpt_family="Alu"
     repeat_region   18004. .18146
                     /rpt_family="Alu"
     repeat_region   18164. .18345
                     /rpt_family="B2"
     repeat_region   18848. .18921
                     /rpt_family="ID"
     repeat_region   18938. .19051
                     /rpt_family="Alu"
     repeat_region   19472. .19645
                     /rpt_family="B2"
     repeat_region   19689. .19795
                     /rpt_family="Alu"
     repeat_region   20808. .20887
                     /rpt_family="L1"
     repeat_region   21309. .21329
                     /rpt_family="B4"
     repeat_region   21330. .21417
                     /rpt_family="Alu"
     repeat_region   21436. .21586
                     /rpt_family="MaLR"
     repeat_region   21557. .21757
                     /rpt_family="MaLR"
     repeat_region   21771. .21812
                     /rpt_family="Alu"
     repeat_region   21948. .22091
                     /rpt_family="Alu"
     repeat_region   22448. .23281
                     /rpt_family="L1"
     repeat_region   23288. .23405
                     /rpt_family="Alu"
     repeat_region   23447. .23515
                     /rpt_family="ID"
     repeat_region   23839. .24523
                     /rpt_family="RMER19A"
     repeat_region   24541. .24659
                     /rpt_family="L1"
     repeat_region   24690. .24764
                     /rpt_family="ID"
     repeat_region   25308. .25445
                     /rpt_family="Alu"
     repeat_region   26207. .26322
                     /rpt_family="Alu"
     repeat_region   27425. .27598
                     /rpt_family="B4"
     repeat_region   27620. .27673
                     /rpt_family="ID"
     repeat_region   27674. .27706
                     /rpt_family="B4"
     repeat_region   27703. .27841
                     /rpt_family="B2"

Alignment Scores:
Pred. No.:                426             Length:          205725
Score:                    9.00            Matches:         9
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              45.00%          Indels:          0
DB:                       10              Gaps:            0

US-10-799-747-116 (1-20) x AC122809 (1-205725)

Qy          3 AlaHisSerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||||||||
Db     136760 GCACATTCAGTTTTGTCATTCTTGCTA 136734

RESULT 10
AC110315
LOCUS       AC110315               231768 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-191N21, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC110315
VERSION     AC110315.5  GI:24818671
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231768)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
```

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 231768)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 231768)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 9, 2002 this sequence version replaced gi:23265946.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GOGP
Center clone name: CH230-191N21
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 218739 bases at least Q40
Consensus quality: 220995 bases at least Q30
Consensus quality: 222405 bases at least Q20
Estimated insert size: 223718; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 230654: contig of 230654 bp in length
* 230655 230754: gap of unknown length
* 230755 231768: contig of 1014 bp in length.

FEATURES Location/Qualifiers
source 1. .231768
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-191N21"
misc_feature 1. .1220
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 1824. .2452
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH282813"
misc_feature complement(222634. .223294)
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH282812"
misc_feature 226573. .227843
/note="wgs_end_extension
clone_end:T7"
misc_feature 228778. .230654
/note="wgs_end_extension
clone_end:T7"
ORIGIN

Alignment Scores:

Pred. No.:	475	Length:	231768
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.00%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AC110315 (1-231768)

```
Qy          6 ValLeuSerPheLeuLeuTrpThrPro 14
              ||||||||||||||||||||||||||
Db     172304 GTGCTTTCCTTCCTGTTATGGACTCCA 172330
```

RESULT 11
AC103176/c
LOCUS AC103176 254839 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-98N18, WORKING DRAFT SEQUENCE, 4 unordered pieces.
ACCESSION AC103176
VERSION AC103176.5 GI:30578547
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 254839)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 254839)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 254839)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23265304.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GJKE
Center clone name: CH230-98N18
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 241531 bases at least Q40
Consensus quality: 243998 bases at least Q30
Consensus quality: 246135 bases at least Q20
Estimated insert size: 252438; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 250539: contig of 250539 bp in length
* 250540 250639: gap of unknown length
* 250640 251762: contig of 1123 bp in length
* 251763 251862: gap of unknown length
* 251863 252935: contig of 1073 bp in length
* 252936 253035: gap of unknown length
* 253036 254839: contig of 1804 bp in length.

FEATURES Location/Qualifiers
source 1. .254839
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-98N18"
misc_feature 1. .1230
/note="wgs_end_extension
clone_end:T7"
misc_feature 3845. .5494
/note="wgs_end_extension
clone_end:T7"
misc_feature complement(6002. .29990)
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH311760"
misc_feature complement(247726. .248554)
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH311788"

ORIGIN

Alignment Scores:

Pred. No.:	518	Length:	254839
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.00%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AC103176 (1-254839)

Qy 5 SerValLeuSerPheLeuLeuTrpThr 13
||||||||||||||||||||||||||
Db 107271 TCTGTACTTAGTTTTCTCCTGTGGACT 107245

RESULT 12
AC098074/c

```
LOCUS       AC098074               267352 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-119I24, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC098074
VERSION     AC098074.7  GI:30522668
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 267352)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 267352)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 267352)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24635703.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGFO
                Center clone name: CH230-119I24
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 254595 bases at least Q40
                Consensus quality: 257284 bases at least Q30
                Consensus quality: 259179 bases at least Q20
                Estimated insert size: 267168; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   242526: contig of 242526 bp in length
            *    242527   242626: gap of unknown length
            *    242627   264340: contig of 21714 bp in length
            *    264341   264440: gap of unknown length
            *    264441   265515: contig of 1075 bp in length
            *    265516   265615: gap of unknown length
            *    265616   267352: contig of 1737 bp in length.
FEATURES             Location/Qualifiers
     source          1. .267352
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-119I24"
     misc_feature    1772. .2630
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH316141"
     misc_feature    18571. .19415
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH316143"
ORIGIN

Alignment Scores:
Pred. No.:                541             Length:             267352
Score:                    9.00            Matches:            9
Percent Similarity:       100.00%         Conservative:       0
Best Local Similarity:    100.00%         Mismatches:         0
Query Match:              45.00%          Indels:             0
```

```
DB:                 2          Gaps:          0

US-10-799-747-116 (1-20) x AC098074 (1-267352)

Qy          5 SerValLeuSerPheLeuLeuTrpThr 13
              ||||||||||||||||||||||||||
Db     247232 TCTGTACTTAGTTTTCTCCTGTGGACT 247206

RESULT 13
AC098226/c
LOCUS       AC098226              306660 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-56K13, *** SEQUENCING IN PROGRESS
            ***, 11 unordered pieces.
ACCESSION   AC098226
VERSION     AC098226.7  GI:30580937
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 306660)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 306660)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 306660)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23268205.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIVQ
                Center clone name: CH230-56K13
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 236654 bases at least Q40
                Consensus quality: 241334 bases at least Q30
                Consensus quality: 244452 bases at least Q20
                Estimated insert size: 240001; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4465: contig of 4465 bp in length
            *     4466     4565: gap of unknown length
            *     4566   280032: contig of 275467 bp in length
            *   280033   280132: gap of unknown length
            *   280133   291663: contig of 11531 bp in length
            *   291664   291763: gap of unknown length
            *   291764   297095: contig of 5332 bp in length
            *   297096   297195: gap of unknown length
            *   297196   298558: contig of 1363 bp in length
            *   298559   298658: gap of unknown length
            *   298659   299805: contig of 1147 bp in length
            *   299806   299905: gap of unknown length
            *   299906   301125: contig of 1220 bp in length
            *   301126   301225: gap of unknown length
            *   301226   302302: contig of 1077 bp in length
            *   302303   302402: gap of unknown length
            *   302403   303604: contig of 1202 bp in length
            *   303605   303704: gap of unknown length
            *   303705   305546: contig of 1842 bp in length
            *   305547   305646: gap of unknown length
            *   305647   306660: contig of 1014 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1. .306660
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-56K13"
     misc_feature    1. .1130
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    4566. .5998
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    29817. .30488
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH272628"
     misc_feature    31694. .33434
                     /note="wgs_contig"
     misc_feature    159947. .161767
                     /note="wgs_contig"
     misc_feature    complement(265403. .266254)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH272626"
     misc_feature    272770. .274704
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    278461. .280032
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

Alignment Scores:
Pred. No.:                613             Length:         306660
Score:                    9.00            Matches:        9
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              45.00%          Indels:         0
DB:                       2               Gaps:           0

US-10-799-747-116 (1-20) x AC098226 (1-306660)

Qy          6 ValLeuSerPheLeuLeuTrpThrPro 14
              |||||||||||||||||||||||||||
Db     215746 GTGCTTTCCTTCCTGTTATGGACTCCA 215720

RESULT 14
G25673
LOCUS       G25673                   339 bp    DNA     linear   STS 02-JUN-1996
DEFINITION  human STS EST113435, sequence tagged site.
ACCESSION   G25673
VERSION     G25673.1  GI:1347905
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: AATTACAGTTTGGGTTCTCAGATT
            Primer B: CCATGCTAAGTCCAGTGGGG
            STS size: 150
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession T59368).
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="527.4 cR from top of Chr2 linkage group"
     STS             1. .150
     primer_bind     1. .24
     primer_bind     complement(131. .150)
ORIGIN

Alignment Scores:
Pred. No.:                13.1            Length:         339
Score:                    8.00            Matches:        8
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              40.00%          Indels:         0
DB:                       11              Gaps:           0

US-10-799-747-116 (1-20) x G25673 (1-339)

Qy          3 AlaHisSerValLeuSerPheLeu 10
              |||||||||||||||||||||||||
Db        108 GCTCATTCTGTACTTTCCTTCCTC 131

RESULT 15
G27758
LOCUS       G27758                   356 bp    DNA     linear   STS 29-JUN-1996
DEFINITION  human STS SHGC-33168, sequence tagged site.
ACCESSION   G27758
VERSION     G27758.1  GI:1396477
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: AATTACAGTTTGGGTTCTCAGATT
            Primer B: CCATGCTAAGTCCAGTGGGG
```

```
          STS size: 150
          PCR Profile:
                  Initial incubation: 94 degrees C for 90 seconds

                  Denaturation:       94 degrees C for 15 seconds
                  Annealing:          62 degrees C for 23 seconds
                  Polymerization:     72 degrees C for 30 seconds
                  PCR Cycles:         30
                  Thermal Cycler:     Perkin Elmer 9600
          Protocol:
                  Template:       25 ng
                  Primer:         each 1 uM
                  dNTPs:          each 200 uM
                  Taq Polymerase: 0.05 units/ul
                  Total Vol:      10 ul

          Buffer:
                  MgCl2:          2.5 mM
                  KCl:            50  mM
                  Tris-HCl:       20  mM
                  pH:             8.3

          Prepared with primer pairs provided by Sandoz, derived from T59368
          -- Washington University/Merck EST sequence.
FEATURES             Location/Qualifiers
     source          1. .356
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="2"
     STS             1. .150
     primer_bind     1. .24
     primer_bind     complement(131. .150)
ORIGIN

Alignment Scores:
Pred. No.:                13.7          Length:          356
Score:                    8.00          Matches:         8
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              40.00%        Indels:          0
DB:                       11            Gaps:            0

US-10-799-747-116 (1-20) x G27758 (1-356)

Qy          3 AlaHisSerValLeuSerPheLeu 10
              ||||||||||||||||||||||||
Db        108 GCTCATTCTGTACTTTCCTTCCTC 131


Search completed: July 21, 2004, 04:43:02
Job time : 1826 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 02:53:21 ; Search time 649 Seconds
(without alignments)
130.915 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 20
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6742599

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161441_21967/app_query.fasta_1.199
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=olig.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747_@CGN_1_1_708_@runat_19072004_161441_21967 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19	95.0	1434	2	AAX37452	Aax37452 Human sec
	2	19	95.0	1434	7	ADA39771	Ada39771 Human sec
	3	19	95.0	1434	7	ACC50424	Acc50424 Human sec
	4	8	40.0	356	6	ABL77790	Abl77790 Human ova
	5	8	40.0	1503	7	ADA70841	Ada70841 Rice gene
c	6	7	35.0	49	7	ACD97100	Acd97100 Human col
	7	7	35.0	65	6	ABZ28939	Abz28939 Candida g
	8	7	35.0	90	4	AAS23646	Aas23646 Tetracycl
	9	7	35.0	109	2	AAH86006	Aah86006 Human sin
	10	7	35.0	155	7	ACD97494	Acd97494 Human col
c	11	7	35.0	263	2	AAT20076	Aat20076 Human gen
	12	7	35.0	264	9	ADC93542	Adc93542 E. faeciu
c	13	7	35.0	293	6	ABN76057	Abn76057 Human ORF
c	14	7	35.0	294	9	ADC92876	Adc92876 E. faeciu
c	15	7	35.0	315	2	AAV88416	Aav88416 EST clone
c	16	7	35.0	352	6	ABQ85808	Abq85808 Arabidops
c	17	7	35.0	363	5	ABV25515	Abv25515 Human pro
c	18	7	35.0	377	5	AAF66743	Aaf66743 Novel hum
c	19	7	35.0	379	8	ACH50509	Ach50509 Human leu
	20	7	35.0	394	4	AAK87798	Aak87798 Human dig
c	21	7	35.0	409	5	ABV38500	Abv38500 Human pro
	22	7	35.0	415	3	AAH30357	Aah30357 Human col
c	23	7	35.0	426	5	ABV08598	Abv08598 Human pro
c	24	7	35.0	432	6	ABK76313	Abk76313 Bacillus
c	25	7	35.0	433	7	ABX61964	Abx61964 Arabidops
	26	7	35.0	436	4	AAK56191	Aak56191 Human imm
c	27	7	35.0	440	8	ACH32772	Ach32772 Human end
c	28	7	35.0	445	8	ACH42384	Ach42384 Human foe
c	29	7	35.0	455	8	ACH15424	Ach15424 Human adu
	30	7	35.0	465	4	AAI11211	Aai11211 Probe #11
	31	7	35.0	465	4	ABA52870	Aba52870 Human foe
	32	7	35.0	465	4	AAI32473	Aai32473 Probe #11
	33	7	35.0	465	4	AAK67520	Aak67520 Human imm
	34	7	35.0	465	4	ABA42442	Aba42442 Human bre
	35	7	35.0	465	4	ABA22654	Aba22654 Probe #11
	36	7	35.0	465	4	AAK26581	Aak26581 Human bon
	37	7	35.0	465	4	AAK01123	Aak01123 Human bra
	38	7	35.0	465	4	ABS26173	Abs26173 Human liv
	39	7	35.0	465	5	AAI01128	Aai01128 Probe #11
	40	7	35.0	465	6	ABS01172	Abs01172 Human gen
	41	7	35.0	466	6	ABN94789	Abn94789 Gene #128
c	42	7	35.0	467	8	ACH26710	Ach26710 Human adu
	43	7	35.0	468	4	AAK32950	Aak32950 Human bon
	44	7	35.0	468	4	AAK07199	Aak07199 Human bra
	45	7	35.0	468	4	ABS32679	Abs32679 Human liv

ALIGNMENTS

RESULT 1
AAX37452
ID AAX37452 standard; cDNA; 1434 BP.
XX
AC AAX37452;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein cDNA fragment containing gene 2.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS Homo sapiens.
XX
PN WO9918208-A1.
XX
PD 15-APR-1999.
XX
PF 01-OCT-1998; 98WO-US020775.
XX
PR 02-OCT-1997; 97US-0060833P.
PR 02-OCT-1997; 97US-0060836P.
PR 02-OCT-1997; 97US-0060837P.
PR 02-OCT-1997; 97US-0060838P.
PR 02-OCT-1997; 97US-0060839P.

PR 02-OCT-1997; 97US-0060843P.
PR 02-OCT-1997; 97US-0060862P.
PR 02-OCT-1997; 97US-0060866P.
PR 02-OCT-1997; 97US-0060874P.
PR 02-OCT-1997; 97US-0060880P.
PR 02-OCT-1997; 97US-0060884P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
PI Lafleur DW, Shi Y;
XX
DR WPI; 1999-264022/22.
DR P-PSDB; AAY07853.
XX
PT New isolated human genes and the secreted polypeptides they encode.
XX
PS Claim 1a; Page 228; 368pp; English.
XX
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in AAY07852-Y07993 and the encoding nucleic acids are
CC represented in AAX37451-X37552
XX
SQ Sequence 1434 BP; 480 A; 204 C; 250 G; 495 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.:	5.49e-10	Length:	1434
Score:	19.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	95.00%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AAX37452 (1-1434)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563
```

RESULT 2
ADA39771
ID ADA39771 standard; cDNA; 1434 BP.
XX
AC ADA39771;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human secreted protein encoding cDNA.
XX
KW Human; secreted protein; cancer; hyperproliferative disorder;
KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
KW anaemia; allergic reaction; asthma; cardiovascular disorder;
KW wound healing; cytostatic; immunosuppressive; nootropic; neuroprotective;
KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
KW vulnerary; cardiant; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO2002102993-A2.
XX
PD 27-DEC-2002.
XX
PF 19-MAR-2002; 2002WO-US008123.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-175238/17.
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating cancer or other hyperproliferative disorder,
PT asthma, allergies or AIDS.
XX
PS Claim 9; SEQ ID NO 153; 3205pp; English.
XX
CC The invention relates to novel genes ADA39629-ADA40565 and proteins
CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.:	5.49e-10	Length:	1434
Score:	19.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	95.00%	Indels:	0
DB:	7	Gaps:	0

US-10-799-747-116 (1-20) x ADA39771 (1-1434)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563
```

RESULT 3
ACC50424
ID ACC50424 standard; cDNA; 1434 BP.

XX
AC ACC50424;
XX
DT 12-JUN-2003 (first entry)
XX
DE Human secreted protein coding sequence, SEQ ID 91.
XX
KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
KW vulnerary; antiinflammatory; nootropic; neuroprotective;
KW antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200295010-A2.
XX
PD 28-NOV-2002.
XX
PF 19-MAR-2002; 2002WO-US009785.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-129429/12.
XX
PT Novel human secreted proteins, useful for detecting, preventing,
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
PT disorders such as arrhythmia.
XX
PS Claim 21; SEQ ID NO 91; 1881pp; English.
XX
CC The present invention relates to novel human secreted proteins (ABR47633-
CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
CC and their coding sequences are useful for the preparation of a diagnostic
CC or pharmaceutical composition for diagnosing or treating a cardiovascular
CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
CC arteriosclerosis and myocardial ischaemia), neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, renal disorders,
CC proliferative disorders and/or cancerous diseases and conditions, for
CC wound healing and epithelial cell proliferation, to treat inflammation or
CC infection, for treating thrombosis and arteriosclerosis, for treating or
CC preventing neural damage which occurs in neuronal disorders or
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
CC disease, to enhance bone and periodontal regeneration and aid in tissue
CC transplants or bone grafts, to prevent skin aging or hair loss, to
CC stimulate growth and differentiation of haematopoietic cells and bone
CC marrow cells when used in combination with other cytokines, to maintain
CC organs before transplantation or for supporting cell culture of primary
CC tissues, to increase or decrease differentiation or proliferation of
CC embryonic stem cells, or to modulate mammalian characteristics or
CC metabolism. Note: The sequence data for this patent was published in
CC electronic format and is available from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.:	5.49e-10	Length:	1434
Score:	19.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	95.00%	Indels:	0
DB:	7	Gaps:	0

US-10-799-747-116 (1-20) x ACC50424 (1-1434)

Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563

RESULT 4
ABL77790
ID ABL77790 standard; cDNA; 356 BP.
XX
AC ABL77790;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:768.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US017756.
XX
PR 26-MAY-2000; 2000US-0207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI; 2002-122075/16.
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX
PS Claim 1; SEQ ID NO 768; 489pp; English.
XX
CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
SQ Sequence 356 BP; 57 A; 84 C; 78 G; 136 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	19.1	Length:	356
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.00%	Indels:	0
DB:	6	Gaps:	0

US-10-799-747-116 (1-20) x ABL77790 (1-356)

Qy 3 AlaHisSerValLeuSerPheLeu 10
|||||||||||||||||||||||||
Db 108 GCTCATTCTGTACTTTCCTTCCTC 131

RESULT 5
ADA70841
ID ADA70841 standard; DNA; 1503 BP.
XX
AC ADA70841;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 4164.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 4164; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 1503 BP; 395 A; 329 C; 383 G; 396 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	72.6	Length:	1503
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.00%	Indels:	0
DB:	7	Gaps:	0

US-10-799-747-116 (1-20) x ADA70841 (1-1503)

```
Qy          3 AlaHisSerValLeuSerPheLeu 10
              ||||||||||||||||||||||||
Db        266 GCACACAGTGTTTTATCTTTTCTT 289
```

RESULT 6
ACD97100/c
ID ACD97100 standard; cDNA; 49 BP.
XX
AC ACD97100;
XX
DT 23-SEP-2003 (first entry)
XX
DE Human colon cancer cell expressed cDNA #5512.
XX
KW Open reading frame detection; genome sequencing; colon cancer;
KW breast cancer; population genome analysis; genetic shift; cancer;
KW antibiotic resistance; antibiotic non-tolerance; congenital disease;
KW agriculture; food crop genome; resistance gene; retrovirus;
KW influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN US2002155438-A1.
XX
PD 24-OCT-2002.
XX
PF 27-SEP-1999; 99US-00406117.
XX
PR 20-NOV-1998; 98US-00196716.
XX
PA (SIMP/) SIMPSON A J G.
PA (NETO/) NETO E D.
PA (BREN/) BRENTANI R R.
XX
PI Simpson AJG, Neto ED, Brentani RR;
XX
DR WPI; 2003-182626/18.
XX
PT Determining open reading frames of genome of an organism e.g. a human
PT suffering from cancer involves use of single oligonucleotide primer at
PT low stringency for preparing single-stranded cDNA from mRNA of
PT individual.
XX
PS Example 9; Page 787; 959pp; English.
XX
CC The invention describes a method of determining open reading frames in
CC the genome of organism, comprising contacting mRNA from cell of organism
CC with a single oligonucleotide primer (I) at low stringency, preparing
CC single-stranded cDNA by reverse transcribing mRNA with (I), amplifying
CC cDNA , sequencing the product, and repeating the contacting, preparing
CC and amplifying steps with different primers and sequencing resulting
CC nucleic acids. The method is useful for: determining that a known
CC nucleotide sequence from a genome of an organism corresponds to a
CC nucleotide sequence of an open reading frame; for preparing a contig,
CC nucleic acid molecule from a genome of an organism; and for sequencing
CC all or part of a genome of an organism. mRNA is obtained from mammalian
CC or human cell which is associated with a pathological condition e.g. a
CC colon cancer or breast cancer cell. The method is useful for analyses of
CC populations of subjects and can be used to carry out genetic analyses of
CC large or small populations. further, it can be used to study living
CC systems to determine if, e.g. there have been genetic shifts which render
CC an individual or population more or less likely to be afflicted with
CC diseases such as cancer, to determine antibiotic resistance or non-
CC tolerance, and so forth. The method can also be used in the study of
CC congenital diseases, and the risk of affliction to a foetus, as well as
CC the study of whether the conditions are likely to be passed to offspring
CC through ova or sperm. The analyses for pathological conditions can be
CC carried out in all animals, plants, birds, fish, etc. Using this method,
CC in the area of agriculture, for example the genomes of food crops can be
CC studied to determine if resistance genes are present, defects in plant
CC genomes can also be studied in this way. Similarly, the method permits
CC determination of the pathogens which integrate into the genome, such as
CC retroviruses and other integrating viruses such as influenza virus, have
CC undergone shifts or mutations, which may require different approaches to
CC therapy. This method is also applied to eukaryotic pathogens, such as
CC trypanosomes, different types of Plasmodium, etc. The method essentially
CC eliminates sequencing of non-coding portions. This sequence represents a
CC polynucleotide isolated from human colon cancer cell cDNA library
XX
SQ Sequence 49 BP; 20 A; 10 C; 10 G; 6 T; 0 U; 3 Other;

Alignment Scores:

Pred. No.:	31.2	Length:	49
Score:	7.00	Matches:	7

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            35.00%         Indels:        0
DB:                     7              Gaps:          0

US-10-799-747-116 (1-20) x ACD97100 (1-49)

Qy          6 ValLeuSerPheLeuLeuTrp 12
              ||||||||||||||||||||||
Db         32 GTTTTAAGCTTCCTCCTTTGG 12

RESULT 7
ABZ28939
ID   ABZ28939 standard; DNA; 65 BP.
XX
AC   ABZ28939;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida gene related tetracyclin promoter PCR primer SEQ ID NO 3022.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal; PCR; primer; ss.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US049486.
XX
PR   29-DEC-2000; 2000US-0259128P.
PR   20-FEB-2001; 2001US-00792024.
PR   22-AUG-2001; 2001US-0314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele of
PT   a gene and placing other allele of the gene under conditional expression.
XX
PS   Claim 76; SEQ ID NO 3022; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of a PCR
CC   primer used in the method of the invention. Note: The sequence data for
CC   this patent is not represented in the printed specification but is based
CC   on sequence information supplied to Derwent by the European Patent Office
XX
SQ   Sequence 65 BP; 17 A; 12 C; 10 G; 26 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              40.5           Length:        65
Score:                  7.00           Matches:       7
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            35.00%         Indels:        0
DB:                     6              Gaps:          0

US-10-799-747-116 (1-20) x ABZ28939 (1-65)

Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db         13 TCTGTATTATCATTCTTATTG 33

RESULT 8
AAS23646
ID   AAS23646 standard; DNA; 90 BP.
XX
AC   AAS23646;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Tetracycline promoter downstream PCR primer (Tet-Down) #21.
XX
KW   Gene identification; essential gene; GRACE; pathogenic fungus;
KW   gene replacement and conditional expression; fungal infection;
KW   PCR primer; Tet-Down; tetracycline promoter; ss.
XX
OS   Escherichia coli.
OS   Candida albicans.
XX
PN   WO200160975-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005551.
XX
PR   18-FEB-2000; 2000US-0183534P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H;
XX
DR   WPI; 2001-489080/53.
XX
PT   Identifying genes essential to fungal metabolisms and identifying
PT   potential therapeutic agents that target these genes.
XX
PS   Disclosure; Page 292; 324pp; English.
XX
CC   The present invention relates to novel methods for constructing fungal
CC   strains useful for identification and validation of gene products as
CC   targets for therapeutic agents, for creating a collection of identified
CC   essential genes, and screening assays for the discovery of new drugs. The
CC   invention provides the GRACE (gene replacement and conditional
CC   expression) method for the construction of mutant organisms referred to
CC   as GRACE strains of the organism. The invention can be applied to any
CC   organism, particularly a pathogenic fungus e.g. Candida albicans,
CC   Aspergillus fumigatus and Cryptococcus neoformans. The methods are useful
CC   to identify agents that may be used in the treatment of fungal
CC   infections. AAS23626-AAS23686 represent tetracycline promoter downstream
CC   PCR primers (Tet-Down) used in the methods of the present invention
XX
SQ   Sequence 90 BP; 23 A; 15 C; 17 G; 35 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              54.7           Length:        90
Score:                  7.00           Matches:       7
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
```

```
Query Match:            35.00%      Indels:        0
DB:                     4           Gaps:          0

US-10-799-747-116 (1-20) x AAS23646 (1-90)

Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db         13 TCTGTATTATCATTCTTATTG 33

RESULT 9
AAH86006
ID   AAH86006 standard; DNA; 109 BP.
XX
AC   AAH86006;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Human single nucleotide polymorphism containing DNA sequence #863.
XX
KW   Biallelic marker; polymorphism; human; disease; diagnosis; treatment;
KW   phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;
KW   transgenic; single nucleotide polymorphism; SNP; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(47,A)
FT                   /*tag= a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9953095-A2.
XX
PD   21-OCT-1999.
XX
PF   30-MAR-1999;   99WO-US006893.
XX
PR   09-APR-1998;   98US-00057871.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Lander ES,  Wang D,  Hudson T;
XX
DR   WPI; 1999-620443/53.
XX
PT   Polymorphic human genomic sequences and related allele-specific probes
PT   and primers, useful for genetic analysis, e.g. diagnosis and monitoring
PT   of disease.
XX
PS   Claim 1; Page 123; 330pp; English.
XX
CC   This invention describes novel human nucleic acid segments (I) containing
CC   polymorphic sites. The polynucleotides of (I) are used for, e.g.
CC   correlating disease polymorphisms (or disease susceptibility) or other
CC   phenotypic traits (e.g. baldness, obesity, fertility, strength, response
CC   to drugs etc.); diagnosing and monitoring e.g. cancer, inflammation,
CC   heart or central nervous system diseases; detecting susceptibility to
CC   microbial infection; treating or preventing such diseases; forensic
CC   analysis; gene therapy; paternity testing; mapping genomic loci
CC   associated with phenotypic traits (and subsequent cloning of the genes
CC   responsible); and the production of transgenic organisms. Antibodies
CC   raised against (I) are useful as diagnostic and therapeutic tools and in
CC   drug screening. AAH85144 - AAH87644 represent the human DNA sequences
CC   containing biallelic polymorphic sites described in the invention
XX
SQ   Sequence 109 BP; 30 A; 28 C; 18 G; 33 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              65.4        Length:        109
Score:                  7.00        Matches:       7
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
Query Match:            35.00%      Indels:        0
DB:                     2           Gaps:          0

US-10-799-747-116 (1-20) x AAH86006 (1-109)

Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db         15 AGTGTACTGTCTTTTCTACTG 35

RESULT 10
ACD97494
ID   ACD97494 standard; cDNA; 155 BP.
XX
AC   ACD97494;
XX
DT   23-SEP-2003  (first entry)
XX
DE   Human colon cancer cell expressed cDNA #5906.
XX
KW   Open reading frame detection; genome sequencing; colon cancer;
KW   breast cancer; population genome analysis; genetic shift; cancer;
KW   antibiotic resistance; antibiotic non-tolerance; congenital disease;
KW   agriculture; food crop genome; resistance gene; retrovirus;
KW   influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   US2002155438-A1.
XX
PD   24-OCT-2002.
XX
PF   27-SEP-1999;   99US-00406117.
XX
PR   20-NOV-1998;   98US-00196716.
XX
PA   (SIMP/) SIMPSON A J G.
PA   (NETO/) NETO E D.
PA   (BREN/) BRENTANI R R.
XX
PI   Simpson AJG,  Neto ED,  Brentani RR;
XX
DR   WPI; 2003-182626/18.
XX
PT   Determining open reading frames of genome of an organism e.g. a human
PT   suffering from cancer involves use of single oligonucleotide primer at
PT   low stringency for preparing single-stranded cDNA from mRNA of
PT   individual.
XX
PS   Example 9; Page 842; 959pp; English.
XX
CC   The invention describes a method of determining open reading frames in
CC   the genome of organism, comprising contacting mRNA from cell of organism
CC   with a single oligonucleotide primer (I) at low stringency, preparing
CC   single-stranded cDNA by reverse transcribing mRNA with (I), amplifying
CC   cDNA , sequencing the product, and repeating the contacting, preparing
CC   and amplifying steps with different primers and sequencing resulting
CC   nucleic acids. The method is useful for: determining that a known
CC   nucleotide sequence from a genome of an organism corresponds to a
CC   nucleotide sequence of an open reading frame; for preparing a contig,
CC   nucleic acid molecule from a genome of an organism; and for sequencing
CC   all or part of a genome of an organism. mRNA is obtained from mammalian
CC   or human cell which is associated with a pathological condition e.g. a
CC   colon cancer or breast cancer cell. The method is useful for analyses of
CC   populations of subjects and can be used to carry out genetic analyses of
CC   large or small populations. further, it can be used to study living
CC   systems to determine if, e.g. there have been genetic shifts which render
CC   an individual or population more or less likely to be afflicted with
CC   diseases such as cancer, to determine antibiotic resistance or non-
CC   tolerance, and so forth. The method can also be used in the study of
CC   congenital diseases, and the risk of affliction to a foetus, as well as
CC   the study of whether the conditions are likely to be passed to offspring
CC   through ova or sperm. The analyses for pathological conditions can be
CC   carried out in all animals, plants, birds, fish, etc. Using this method,
CC   in the area of agriculture, for example the genomes of food crops can be
```

CC studied to determine if resistance genes are present, defects in plant
CC genomes can also be studied in this way. Similarly, the method permits
CC determination of the pathogens which integrate into the genome, such as
CC retroviruses and other integrating viruses such as influenza virus, have
CC undergone shifts or mutations, which may require different approaches to
CC therapy. This method is also applied to eukaryotic pathogens, such as
CC trypanosomes, different types of Plasmodium, etc. The method essentially
CC eliminates sequencing of non-coding portions. This sequence represents a
CC polynucleotide isolated from human colon cancer cell cDNA library
XX
SQ Sequence 155 BP; 41 A; 30 C; 33 G; 49 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.:	90.6	Length:	155
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	35.00%	Indels:	0
DB:	7	Gaps:	0

US-10-799-747-116 (1-20) x ACD97494 (1-155)

```
Qy          5 SerValLeuSerPheLeuLeu 11
              ||||||||||||||||||||
Db         12 AGTGTTTTGTCTTTTCTCTTA 32
```

RESULT 11
AAT20076/c
ID AAT20076 standard; cDNA to mRNA; 263 BP.
XX
AC AAT20076;
XX
DT 17-JUL-1996 (first entry)
XX
DE Human gene signature HUMGS01218.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP001916.
XX
PR 12-NOV-1993; 93JP-00355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206931/27.
XX
PT Single-stranded DNA for identifying gene signatures - isolated from 3'-
PT directed human cDNA library that reflects relative abundance of corresp.
PT mRNA in specific human tissues.
XX
PS Claim 1; Page 554; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp. double
CC -stranded DNA) which comprises one of the 7837 "GS" sequences given in
CC AAT19001-T26837 and which is able to hybridise to part of human genomic
CC DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were
CC obtained from 3'-directed cDNA libraries prepared from various human
CC tissues; synthesis of cDNA was initiated from the 3'-end of mRNA by using
CC poly(T) as the sole primer. Since the 3'- untranslated sequence is unique
CC to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise
CC with specific mRNAs. Each library is constructed so as to reflect
CC accurately the relative abundance of different mRNAs in the particular
CC tissue from which it was derived. The appearance frequency of a given GS
CC in a cDNA library can be determined (esp. using primers and probes
CC derived from the GS sequences) as a means of diagnosing abnormal cell
CC function or for recognising different cell types
XX
SQ Sequence 263 BP; 92 A; 34 C; 43 G; 92 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.:	148	Length:	263
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	35.00%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AAT20076 (1-263)

```
Qy          5 SerValLeuSerPheLeuLeu 11
              ||||||||||||||||||||
Db         82 AGTGTGCTTTCTTTTCTCCTC 62
```

RESULT 12
ADC93542
ID ADC93542 standard; DNA; 264 BP.
XX
AC ADC93542;
XX
DT 01-JAN-2004 (first entry)
XX
DE E. faecium DNA sequence SEQ ID 3169.
XX
KW ds; gene; urinary tract infection; bacteraemia; endocarditis; wound;
KW abdominal-pelvic infection.
XX
OS Enterococcus faecium.
XX
PN US6583275-B1.
XX
PD 24-JUN-2003.
XX
PF 30-JUN-1998; 98US-00107532.
XX
PR 02-JUL-1997; 97US-0051571P.
PR 14-MAY-1998; 98US-0085598P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2003-799836/75.
DR P-PSDB; ADC97196.
XX
PT New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
PS Example 1; SEQ ID NO 3169; 243pp; English.
XX
CC The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids is useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -

CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium nucleic acids.
XX
SQ Sequence 264 BP; 67 A; 59 C; 48 G; 88 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.:	148	Length:	264
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	35.00%	Indels:	0
DB:	9	Gaps:	0

US-10-799-747-116 (1-20) x ADC93542 (1-264)

```
Qy          5 SerValLeuSerPheLeuLeu 11
              ||||||||||||||||||||
Db        104 TCGGTTCTGTCTTTCCTTTTA 124
```

RESULT 13
ABN76057/c
ID ABN76057 standard; cDNA; 293 BP.
XX
AC ABN76057;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF1004 cDNA, SEQ ID NO:2007.
XX
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200190366-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US017076.
XX
PR 24-MAY-2000; 2000US-0206690P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Leach MD, Shimkets RA;
XX
DR WPI; 2002-106200/14.
DR P-PSDB; ABP32031.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.
XX
PS Claim 1; Page 766; 2508pp; English.
XX
CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising CRFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases
XX
SQ Sequence 293 BP; 74 A; 85 C; 66 G; 66 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.:	163	Length:	293
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	35.00%	Indels:	0
DB:	6	Gaps:	0

US-10-799-747-116 (1-20) x ABN76057 (1-293)

```
Qy          5 SerValLeuSerPheLeuLeu 11
              ||||||||||||||||||||
Db        158 AGTGTTCTCAGCTTTCTCCTT 138
```

RESULT 14
ADC92876/c
ID ADC92876 standard; DNA; 294 BP.
XX
AC ADC92876;
XX
DT 01-JAN-2004 (first entry)
XX
DE E. faecium DNA sequence SEQ ID 2503.
XX
KW ds; gene; urinary tract infection; bacteraemia; endocarditis; wound;
KW abdominal-pelvic infection.
XX
OS Enterococcus faecium.
XX
PN US6583275-B1.
XX
PD 24-JUN-2003.
XX
PF 30-JUN-1998; 98US-00107532.
XX
PR 02-JUL-1997; 97US-0051571P.
PR 14-MAY-1998; 98US-0085598P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX

PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2003-799836/75.
DR P-PSDB; ADC96530.
XX
PT New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
PS Example 1; SEQ ID NO 2503; 243pp; English.
XX
CC The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids is useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -
CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium nucleic acids.
XX
SQ Sequence 294 BP; 106 A; 57 C; 61 G; 68 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.:	164	Length:	294
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	35.00%	Indels:	0
DB:	9	Gaps:	0

US-10-799-747-116 (1-20) x ADC92876 (1-294)

```
Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db         64 TCGGTTCTGTCTTTCCTTTTA 44
```

RESULT 15
AAV88416/c
ID AAV88416 standard; cDNA; 315 BP.
XX
AC AAV88416;
XX
DT 12-FEB-1999 (first entry)
XX
DE EST clone EK581.
XX
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS Homo sapiens.
XX
PN WO9845437-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US006956.
XX
PR 10-APR-1997; 97US-00837312.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, Mccoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
PI Spaulding V, Agostino MJ;
XX
DR WPI; 1999-070078/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from e.g.
PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,
PT pituitary, retina and colon cDNA libraries.
XX
PS Claim 1; Page 389; 641pp; English.
XX
CC The present sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene therapy
XX
SQ Sequence 315 BP; 74 A; 79 C; 65 G; 97 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	175	Length:	315
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	35.00%	Indels:	0
DB:	2	Gaps:	0

US-10-799-747-116 (1-20) x AAV88416 (1-315)

```
Qy          2 AlaAlaHisSerValLeuSer 8
              ||||||||||||||||||||
Db        215 GCAGCACACAGTGTCCTTTCT 195
```

Search completed: July 21, 2004, 04:12:23
Job time : 660 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 03:55:26 ; Search time 72 Seconds
(without alignments)
154.153 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 20
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1359535

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161443_22019/app_query.fasta_1.199
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=olig.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747_@CGN_1_1_105_@runat_19072004_161443_22019 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	7	35.0	264	4	US-09-107-532A-3169	Sequence 3169, Ap
c	2	7	35.0	294	4	US-09-107-532A-2503	Sequence 2503, Ap
c	3	7	35.0	894	4	US-09-107-532A-1065	Sequence 1065, Ap
c	4	7	35.0	933	4	US-09-308-003-46	Sequence 46, Appl
	5	7	35.0	1983	1	US-08-221-817-21	Sequence 21, Appl
	6	7	35.0	1983	1	US-08-454-439-21	Sequence 21, Appl
	7	7	35.0	1983	5	PCT-US94-10487-21	Sequence 21, Appl
	8	7	35.0	2056	2	US-08-836-442-1	Sequence 1, Appli
	9	7	35.0	2560	4	US-09-786-256C-29	Sequence 29, Appl
	10	7	35.0	2569	4	US-08-956-171E-74	Sequence 74, Appl
c	11	7	35.0	3024	2	US-08-836-943-1	Sequence 1, Appli
c	12	7	35.0	3465	4	US-09-134-000C-2717	Sequence 2717, Ap
c	13	7	35.0	3892	3	US-08-569-214-1	Sequence 1, Appli
c	14	7	35.0	3892	3	US-08-937-236-1	Sequence 1, Appli
c	15	7	35.0	4218	4	US-09-081-385-8	Sequence 8, Appli
	16	7	35.0	34185	4	US-09-545-481-3	Sequence 3, Appli
c	17	7	35.0	99916	4	US-09-816-095-3	Sequence 3, Appli
c	18	7	35.0	536165	4	US-09-214-808-1	Sequence 1, Appli
c	19	7	35.0	786431	4	US-09-751-389-3	Sequence 3, Appli
	20	7	35.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	21	7	35.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	22	6	30.0	18	4	US-09-308-003-40	Sequence 40, Appl
c	23	6	30.0	18	4	US-09-422-978-9787	Sequence 9787, Ap
	24	6	30.0	20	1	US-08-271-942A-86	Sequence 86, Appl
	25	6	30.0	20	3	US-08-779-916A-86	Sequence 86, Appl
	26	6	30.0	20	5	PCT-US95-08604-86	Sequence 86, Appl
	27	6	30.0	29	3	US-08-544-381B-152	Sequence 152, App
c	28	6	30.0	32	3	US-08-575-967A-6	Sequence 6, Appli
c	29	6	30.0	35	3	US-09-282-736-16	Sequence 16, Appl
	30	6	30.0	40	4	US-10-092-246-28	Sequence 28, Appl
	31	6	30.0	40	4	US-10-150-051-7	Sequence 7, Appli
	32	6	30.0	40	4	US-10-150-051-8	Sequence 8, Appli
c	33	6	30.0	45	4	US-08-983-564A-21	Sequence 21, Appl
	34	6	30.0	47	4	US-09-422-978-2157	Sequence 2157, Ap
c	35	6	30.0	65	2	US-09-121-887-5	Sequence 5, Appli
c	36	6	30.0	65	3	US-09-241-353-5	Sequence 5, Appli
c	37	6	30.0	65	3	US-09-245-984-5	Sequence 5, Appli
c	38	6	30.0	65	3	US-09-241-979-5	Sequence 5, Appli
c	39	6	30.0	107	4	US-09-313-294A-1806	Sequence 1806, Ap
c	40	6	30.0	108	1	US-08-441-591-25	Sequence 25, Appl
c	41	6	30.0	108	1	US-08-303-362A-25	Sequence 25, Appl
c	42	6	30.0	108	5	PCT-US95-05600-42	Sequence 42, Appl
c	43	6	30.0	117	3	US-09-284-832-33	Sequence 33, Appl
c	44	6	30.0	120	3	US-09-284-832-32	Sequence 32, Appl
	45	6	30.0	138	3	US-09-432-020B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-107-532A-3169
; Sequence 3169, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277

```
;   INFORMATION FOR SEQ ID NO: 3169:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 264 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...264
;        SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-532A-3169

Alignment Scores:
Pred. No.:                15              Length:         264
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x US-09-107-532A-3169 (1-264)

Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db        104 TCGGTTCTGTCTTTCCTTTTA 124

RESULT 2
US-09-107-532A-2503/c
; Sequence 2503, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 2503:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 294 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...294
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2503:
US-09-107-532A-2503

Alignment Scores:
Pred. No.:                16.6            Length:         294
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x US-09-107-532A-2503 (1-294)

Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db         64 TCGGTTCTGTCTTTCCTTTTA 44

RESULT 3
US-09-107-532A-1065/c
; Sequence 1065, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 1065:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 894 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
```

```
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...894
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1065:
US-09-107-532A-1065

Alignment Scores:
Pred. No.:                48              Length:         894
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x US-09-107-532A-1065 (1-894)

Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db        887 TCTGTTCTTAGCTTTTTACTC 867

RESULT 4
US-09-308-003-46/c
; Sequence 46, Application US/09308003
; Patent No. 6326170
; GENERAL INFORMATION:
;  APPLICANT: Burnham, Martin K. R.
;  APPLICANT: Lonetto, Michael A.
;  APPLICANT: Warren, Patrick V.
;  TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,
;  TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES
;  FILE REFERENCE: GM10093
;  CURRENT APPLICATION NUMBER: US/09/308,003
;  CURRENT FILING DATE: 1999-05-10
;  EARLIER APPLICATION NUMBER: 60/058,710
;  EARLIER FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 46
;   LENGTH: 933
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(933)
US-09-308-003-46

Alignment Scores:
Pred. No.:                50              Length:         933
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x US-09-308-003-46 (1-933)

Qy          4 HisSerValLeuSerPheLeu 10
              |||||||||||||||||||||
Db        891 CATTCTGTTCTCAGCTTCTTG 871

RESULT 5
US-08-221-817-21
; Sequence 21, Application US/08221817
; Patent No. 5532151
;  GENERAL INFORMATION:
;    APPLICANT:  Chantry, David
;    APPLICANT:  Gray, Patrick W.
;    APPLICANT:  Hoekstra, Merle F.
;    TITLE OF INVENTION:  A No. 5532151el G Protein-Coupled Receptor
;    TITLE OF INVENTION:  Kinase GRK6
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/221,817
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/123,932
;      FILING DATE:  17 SEP 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5532151and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31981
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1983 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  13..1740
US-08-221-817-21

Alignment Scores:
Pred. No.:                103             Length:         1983
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       1               Gaps:           0

US-10-799-747-116 (1-20) x US-08-221-817-21 (1-1983)

Qy          3 AlaHisSerValLeuSerPhe 9
              |||||||||||||||||||
Db       1262 GCTCACTCTGTTCTCAGCTTC 1282

RESULT 6
US-08-454-439-21
; Sequence 21, Application US/08454439
; Patent No. 5591618
;  GENERAL INFORMATION:
;    APPLICANT:  Chantry, David
;    APPLICANT:  Gray, Patrick W.
;    APPLICANT:  Hoekstra, Merle F.
;    TITLE OF INVENTION:  A No. 5591618el G Protein-Coupled Receptor
;    TITLE OF INVENTION:  Kinase GRK6
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/454,439
;       FILING DATE:  30-MAY-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/221,817
;       FILING DATE:  31-MAR-1994
;       APPLICATION NUMBER:  08/123,932
;       FILING DATE:  17 SEP 1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 5591618and, Greta E.
;       REGISTRATION NUMBER:  35,302
;       REFERENCE/DOCKET NUMBER:  31981
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (312) 474-6300
;       TELEFAX:  (312) 474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1983 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  13..1740
US-08-454-439-21

Alignment Scores:
Pred. No.:                103             Length:         1983
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       1               Gaps:           0

US-10-799-747-116 (1-20) x US-08-454-439-21 (1-1983)

Qy          3 AlaHisSerValLeuSerPhe 9
              |||||||||||||||||||
Db       1262 GCTCACTCTGTTCTCAGCTTC 1282

RESULT 7
PCT-US94-10487-21
; Sequence 21, Application PC/TUS9410487
;  GENERAL INFORMATION:
;     APPLICANT:  ICOS Corporation
;     TITLE OF INVENTION:  A Novel G Protein-Coupled Receptor
;     TITLE OF INVENTION:  Kinase GRK6
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;       ADDRESSEE:  Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  USA
;       ZIP:  60606
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/10487
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/221,817
;       FILING DATE:  31 MAR 1994
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/123,932
;       FILING DATE:  17 SEP 1993
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Noland, Greta E.
;       REGISTRATION NUMBER:  35,302
;       REFERENCE/DOCKET NUMBER:  27866/31981
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (312) 474-6300
;       TELEFAX:  (312) 474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1983 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  13..1740
PCT-US94-10487-21

Alignment Scores:
Pred. No.:                103             Length:         1983
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       5               Gaps:           0

US-10-799-747-116 (1-20) x PCT-US94-10487-21 (1-1983)

Qy          3 AlaHisSerValLeuSerPhe 9
              |||||||||||||||||||
Db       1262 GCTCACTCTGTTCTCAGCTTC 1282

RESULT 8
US-08-836-442-1
; Sequence 1, Application US/08836442
; Patent No. 5990293
;  GENERAL INFORMATION:
;     APPLICANT:  DOCHERTY, Andrew, J.P.
;     APPLICANT:  SLOCOMBE, Patrick, M.
;     TITLE OF INVENTION:  A HUMAN METALLOPROTEINASE
;     TITLE OF INVENTION:  VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;       STREET:  130 Water Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/836,442
;       FILING DATE:  01-MAY-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/GB96/02181
;       FILING DATE:  13-MAR-1997
;       APPLICATION NUMBER:  GB 9612150.4
;       FILING DATE:  11-JUN-1996
;       APPLICATION NUMBER:  GB 9526229.1
;       FILING DATE:  21-DEC-1995
```

```
;       APPLICATION NUMBER:  GB 9521498.7
;       FILING DATE:  20-OCT-1995
;       APPLICATION NUMBER:  GB 95521495.3
;       FILING DATE:  20-OCT-1995
;       APPLICATION NUMBER:  GB 9518023.8
;       FILING DATE:  05-SEP-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Resnick, David S
;       REGISTRATION NUMBER:  34,235
;       REFERENCE/DOCKET NUMBER:  47425
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-523-3400
;       TELEFAX:  617-523-6440
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2056 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-836-442-1

Alignment Scores:
Pred. No.:                107             Length:         2056
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       2               Gaps:           0

US-10-799-747-116 (1-20) x US-08-836-442-1 (1-2056)

Qy          4 HisSerValLeuSerPheLeu 10
              ||||||||||||||||||||
Db       1656 CATTCTGTCCTATCCTTCTTA 1676

RESULT 9
US-09-786-256C-29
; Sequence 29, Application US/09786256C
; Patent No. 6680189
; GENERAL INFORMATION:
;  APPLICANT: YOSHIMURA, Koji
;  APPLICANT:  HIKICHI, Yuichi
;  APPLICANT:  NISHIMURA, Atsushi
;  TITLE OF INVENTION: No. 6680189el Protein and DNA Thereof
;  FILE REFERENCE: 2544 US0P
;  CURRENT APPLICATION NUMBER: US/09/786,256C
;  CURRENT FILING DATE:  2001-03-02
;  PRIOR APPLICATION NUMBER: PCT/JP99/04766
;  PRIOR FILING DATE: 1999-09-02
;  PRIOR APPLICATION NUMBER: JP 10-250115
;  PRIOR FILING DATE: 1998-09-03
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
;   LENGTH: 2560
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (55)..(1674)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2560)
;   OTHER INFORMATION: DNA sequence of FIG 1-2 containing SEQ ID NO:3 encoding for prote
;   OTHER INFORMATION: in of SEQ ID NO: 1
;   FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: (55)..()
US-09-786-256C-29

Alignment Scores:
Pred. No.:                131             Length:         2560
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x US-09-786-256C-29 (1-2560)

Qy          4 HisSerValLeuSerPheLeu 10
              ||||||||||||||||||||
Db       1703 CATTCTGTCCTATCCTTCTTA 1723

RESULT 10
US-08-956-171E-74
; Sequence 74, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;       APPLICANT: Charles Kunsch
;                  Gil H. Choi
;                  Patrick S. Dillon
;                  Craig A. Rosen
;                  Steven C. Barash
;                  Michael R. Fannon
;       TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;       NUMBER OF SEQUENCES: 5256
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;            COMPUTER: HP Vectra 486/33
;            OPERATING SYSTEM: MSDOS version 6.2
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/956,171E
;            FILING DATE: 20-Oct-1997
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/009,861
;            FILING DATE: January 5, 1996
;            APPLICATION NUMBER: 08/781,986
;            FILING DATE: January 3, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mark J. Hyman
;            REGISTRATION NUMBER: 46,789
;            REFERENCE/DOCKET NUMBER: PB248P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (240) 314-1224
;            TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 74:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2569 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-08-956-171E-74

Alignment Scores:
Pred. No.:                132             Length:         2569
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x US-08-956-171E-74 (1-2569)

Qy          4 HisSerValLeuSerPheLeu 10
```

```
                    ||||||||||||||||||||
Db       1609 CATTCTGTTCTCAGCTTCTTG 1629

RESULT 11
US-08-836-943-1/c
; Sequence 1, Application US/08836943
; Patent No. 5965391
;  GENERAL INFORMATION:
;    APPLICANT:  Reinscheid, Dieter
;    APPLICANT:  Eikmanns, Bernhard
;    APPLICANT:  Sahm, Hermann
;    TITLE OF INVENTION:  DNA WHICH REGULATES GENE EXPRESSION IN
;    TITLE OF INVENTION:  CORYNEFORM BACTERIA
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Firm of Karl F. Ross, PC
;      STREET:  5676 Riverdale Ave.
;      CITY:  Bronx
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10471
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,943
;      FILING DATE:  08-MAY-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Jonathan
;      REGISTRATION NUMBER:  26,963
;      REFERENCE/DOCKET NUMBER:  20357
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (718) 884-6600
;      TELEFAX:  718/601-1099
;      TELEX:  620428
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3024 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Corynebacterium glutamicum
;      STRAIN:  ATCC 13032
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  586..2805
US-08-836-943-1

Alignment Scores:
Pred. No.:                154             Length:         3024
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       2               Gaps:           0

US-10-799-747-116 (1-20) x US-08-836-943-1 (1-3024)

Qy          5 SerValLeuSerPheLeuLeu 11
                ||||||||||||||||||||
Db       2812 AGCGTGCTTAGTTTTTTGCTT 2792

RESULT 12
US-09-134-000C-2717/c
; Sequence 2717, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2717
;   LENGTH: 3465
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-2717

Alignment Scores:
Pred. No.:                176             Length:         3465
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x US-09-134-000C-2717 (1-3465)

Qy          5 SerValLeuSerPheLeuLeu 11
                ||||||||||||||||||||
Db         32 TCTGTTCTCTCCTTTCTTTTA 12

RESULT 13
US-08-569-214-1/c
; Sequence 1, Application US/08569214
; Patent No. 6165469
;  GENERAL INFORMATION:
;    APPLICANT:  MANN, BARBARA J.
;    APPLICANT:  PETRI, WILLIAM A.
;    TITLE OF INVENTION:  RECOMBINANT ENTAMOEBA HISTOLYTICA LECTIN
;    TITLE OF INVENTION:  SUBUNIT PEPTIDES AND REAGANTS SPECIFIC FOR MEMBERS OF THE
;    TITLE OF INVENTION:  170 KD SUBUNIT MULTIGENE FAMILY
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 PENNSYLVANIA AVENUE N.W., STE. 5500
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  200006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/569,214
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/06890
;      FILING DATE:  17-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  9148-0006.21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3892 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..3873, 3877..3882, 3886..3891)
US-08-569-214-1

Alignment Scores:
Pred. No.:                196             Length:         3892
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       3               Gaps:           0

US-10-799-747-116 (1-20) x US-08-569-214-1 (1-3892)

Qy          4 HisSerValLeuSerPheLeu 10
              ||||||||||||||||||||
Db        230 CATTCTGTATTATCATTTTTA 210

RESULT 14
US-08-937-236-1/c
; Sequence 1, Application US/08937236
; Patent No. 6187310
;  GENERAL INFORMATION:
;    APPLICANT:  MANN, BARBARA J.
;    APPLICANT:  PETRI, WILLIAM A.
;    APPLICANT:  DODSON, JAMES M.
;    TITLE OF INVENTION:  RECOMBINANT ENTAMOEBA HISTOLYTICA LECTIN
;    TITLE OF INVENTION:  SUBUNIT PEPTIDES AND REAGANTS SPECIFIC FOR MEMBERS OF THE
;    TITLE OF INVENTION:  170 KD SUBUNIT MULTIGENE FAMILY
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 PENNSYLVANIA AVENUE N.W., STE. 5500
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  200006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/937,236
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/569,214
;      FILING DATE:  16 SEPTEMBER 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LIVNAT, SHMUEL
;      REGISTRATION NUMBER:  33,949
;      REFERENCE/DOCKET NUMBER:  291482000622
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3892 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..3873, 3877..3882, 3886..3891)
US-08-937-236-1

Alignment Scores:
Pred. No.:                196             Length:         3892
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       3               Gaps:           0

US-10-799-747-116 (1-20) x US-08-937-236-1 (1-3892)

Qy          4 HisSerValLeuSerPheLeu 10
              ||||||||||||||||||||
Db        230 CATTCTGTATTATCATTTTTA 210

RESULT 15
US-09-081-385-8/c
; Sequence 8, Application US/09081385
; Patent No. 6593456
;  GENERAL INFORMATION:
;    APPLICANT:  Gatanaga, T.
;    APPLICANT:  Granger, G.A.
;    TITLE OF INVENTION:  Factors Altering Tumor Necrosis
;    TITLE OF INVENTION:  Factor Receptor Releasing Enzyme Activity, and Methods
;    TITLE OF INVENTION:  of Use Thereof
;    NUMBER OF SEQUENCES:  154
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/081,385
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/964,747
;      FILING DATE:  05-NOV-1997
;      APPLICATION NUMBER:  60/030,761
;      FILING DATE:  06-NOV-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wu, Frank
;      REGISTRATION NUMBER:  41,386
;      REFERENCE/DOCKET NUMBER:  22000-20577.21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-081-385-8

Alignment Scores:
Pred. No.:                212             Length:         4218
Score:                    7.00            Matches:        7
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              35.00%          Indels:         0
DB:                       4               Gaps:           0

US-10-799-747-116 (1-20) x US-09-081-385-8 (1-4218)

Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||
```

Db 449 TCAGTGTTATCCTTCCTCTTA 429

Search completed: July 21, 2004, 05:24:13
Job time : 94 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 04:43:08 ; Search time 3357 Seconds
(without alignments)
29.065 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 20
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3191023 seqs, 2439312756 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6371784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161446_22092/app_query.fasta_1.199
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=olig.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10799747_@CGN_1_1_723_@runat_19072004_161446_22092
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19	95.0	1434	15	US-10-195-730-12	Sequence 12, Appl
	2	8	40.0	356	9	US-09-867-701-768	Sequence 768, App
c	3	8	40.0	670	13	US-10-027-632-187894	Sequence 187894,
c	4	8	40.0	670	16	US-10-027-632-187894	Sequence 187894,
c	5	8	40.0	740	13	US-10-027-632-11891	Sequence 11891, A
c	6	8	40.0	740	16	US-10-027-632-11891	Sequence 11891, A
	7	8	40.0	1627	17	US-10-437-963-82948	Sequence 82948, A
	8	8	40.0	2037	17	US-10-437-963-82916	Sequence 82916, A
	9	8	40.0	66681	17	US-10-322-281-411	Sequence 411, App
c	10	8	40.0	133787	17	US-10-322-281-858	Sequence 858, App
	11	8	40.0	196063	17	US-10-322-281-612	Sequence 612, App
c	12	8	40.0	1601042	13	US-10-027-632-59064	Sequence 59064, A
c	13	8	40.0	1601042	16	US-10-027-632-59064	Sequence 59064, A
	14	7	35.0	65	15	US-10-032-585-3022	Sequence 3022, Ap
	15	7	35.0	163	13	US-10-424-599-127928	Sequence 127928,
c	16	7	35.0	191	13	US-10-424-599-98581	Sequence 98581, A
	17	7	35.0	221	13	US-10-085-783A-36161	Sequence 36161, A
	18	7	35.0	221	16	US-10-242-535A-36161	Sequence 36161, A
	19	7	35.0	240	13	US-10-424-599-132181	Sequence 132181,
c	20	7	35.0	279	13	US-10-424-599-22323	Sequence 22323, A
	21	7	35.0	292	13	US-10-424-599-133734	Sequence 133734,
c	22	7	35.0	293	11	US-09-864-408A-2007	Sequence 2007, Ap
c	23	7	35.0	315	14	US-10-040-739-894	Sequence 894, App
	24	7	35.0	324	13	US-10-027-632-143650	Sequence 143650,
	25	7	35.0	324	16	US-10-027-632-143650	Sequence 143650,
	26	7	35.0	348	17	US-10-437-963-77562	Sequence 77562, A
c	27	7	35.0	352	9	US-09-770-791-678	Sequence 678, App
c	28	7	35.0	379	10	US-09-918-995-37721	Sequence 37721, A
c	29	7	35.0	415	17	US-10-437-963-38874	Sequence 38874, A
c	30	7	35.0	423	13	US-10-027-632-104598	Sequence 104598,
	31	7	35.0	423	13	US-10-027-632-298016	Sequence 298016,
c	32	7	35.0	423	16	US-10-027-632-104598	Sequence 104598,
	33	7	35.0	423	16	US-10-027-632-298016	Sequence 298016,
c	34	7	35.0	431	17	US-10-437-963-44015	Sequence 44015, A
c	35	7	35.0	432	9	US-09-974-300-3604	Sequence 3604, Ap
c	36	7	35.0	433	13	US-09-770-423-79	Sequence 79, Appl
c	37	7	35.0	435	13	US-10-424-599-41140	Sequence 41140, A
c	38	7	35.0	440	10	US-09-918-995-19984	Sequence 19984, A
c	39	7	35.0	442	13	US-10-424-599-59524	Sequence 59524, A
c	40	7	35.0	445	10	US-09-918-995-29596	Sequence 29596, A
c	41	7	35.0	455	10	US-09-918-995-2636	Sequence 2636, Ap
	42	7	35.0	455	13	US-10-027-632-179616	Sequence 179616,
	43	7	35.0	455	16	US-10-027-632-179616	Sequence 179616,
	44	7	35.0	465	9	US-09-864-761-1120	Sequence 1120, Ap
c	45	7	35.0	465	9	US-09-796-692-7472	Sequence 7472, Ap

ALIGNMENTS

RESULT 1
US-10-195-730-12
; Sequence 12, Application US/10195730
; Publication No. US20030144492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: PZ017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-195-730-12

```
Alignment Scores:
Pred. No.:                1.22e-10        Length:          1434
Score:                    19.00           Matches:         19
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              95.00%          Indels:          0
DB:                       15              Gaps:            0

US-10-799-747-116 (1-20) x US-10-195-730-12 (1-1434)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 563

RESULT 2
US-09-867-701-768
; Sequence 768, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 768
;   LENGTH: 356
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(356)
;   OTHER INFORMATION: n = A,T,C or G
US-09-867-701-768

Alignment Scores:
Pred. No.:                11.2            Length:          356
Score:                    8.00            Matches:         8
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              40.00%          Indels:          0
DB:                       9               Gaps:            0

US-10-799-747-116 (1-20) x US-09-867-701-768 (1-356)

Qy          3 AlaHisSerValLeuSerPheLeu 10
              |||||||||||||||||||||||||
Db        108 GCTCATTCTGTACTTTCCTTCCTC 131

RESULT 3
US-10-027-632-187894/c
; Sequence 187894, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 187894
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(670)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-187894

Alignment Scores:
Pred. No.:                20              Length:          670
Score:                    8.00            Matches:         8
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              40.00%          Indels:          0
DB:                       13              Gaps:            0

US-10-799-747-116 (1-20) x US-10-027-632-187894 (1-670)

Qy          1 MetAlaAlaHisSerValLeuSer 8
              ||||||||||||||||||||||||
Db         36 ATGGCAGCCCATAGTGTGCTCTCC 13

RESULT 4
US-10-027-632-187894/c
; Sequence 187894, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 187894
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(670)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-187894

Alignment Scores:
Pred. No.:                20              Length:          670
Score:                    8.00            Matches:         8
Percent Similarity:       100.00%         Conservative:    0
```

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            40.00%         Indels:        0
DB:                     16             Gaps:          0

US-10-799-747-116 (1-20) x US-10-027-632-187894 (1-670)

Qy          1 MetAlaAlaHisSerValLeuSer 8
              ||||||||||||||||||||||||
Db         36 ATGGCAGCCCATAGTGTGCTCTCC 13

RESULT 5
US-10-027-632-11891/c
; Sequence 11891, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11891
;   LENGTH: 740
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-11891

Alignment Scores:
Pred. No.:              21.9           Length:        740
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            40.00%         Indels:        0
DB:                     13             Gaps:          0

US-10-799-747-116 (1-20) x US-10-027-632-11891 (1-740)

Qy          5 SerValLeuSerPheLeuLeuTrp 12
              ||||||||||||||||||||||||
Db        266 AGTGTCTTAAGCTTCCTTCTCTGG 243

RESULT 6
US-10-027-632-11891/c
; Sequence 11891, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11891
;   LENGTH: 740
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-11891

Alignment Scores:
Pred. No.:              21.9           Length:        740
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            40.00%         Indels:        0
DB:                     16             Gaps:          0

US-10-799-747-116 (1-20) x US-10-027-632-11891 (1-740)

Qy          5 SerValLeuSerPheLeuLeuTrp 12
              ||||||||||||||||||||||||
Db        266 AGTGTCTTAAGCTTCCTTCTCTGG 243

RESULT 7
US-10-437-963-82948
; Sequence 82948, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 82948
;   LENGTH: 1627
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_82327C.1
US-10-437-963-82948

Alignment Scores:
Pred. No.:              45.2           Length:        1627
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            40.00%         Indels:        0
DB:                     17             Gaps:          0

US-10-799-747-116 (1-20) x US-10-437-963-82948 (1-1627)

Qy          3 AlaHisSerValLeuSerPheLeu 10
              ||||||||||||||||||||||||
Db        266 GCACACAGTGTTTTATCTTTTCTT 289

RESULT 8
```

US-10-437-963-82916
; Sequence 82916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 82916
; LENGTH: 2037
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_82299C.1
US-10-437-963-82916

Alignment Scores:

Pred. No.:	55.6	Length:	2037
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.00%	Indels:	0
DB:	17	Gaps:	0

US-10-799-747-116 (1-20) x US-10-437-963-82916 (1-2037)

```
Qy          6 ValLeuSerPheLeuLeuTrpThr 13
              |||||||||||||||||||||||
Db         69 GTGCTATCCTTTCTCCTGTGGACA 92
```

RESULT 9
US-10-322-281-411
; Sequence 411, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 411
; LENGTH: 66681
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(66681)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-411

Alignment Scores:

Pred. No.:	1.37e+03	Length:	66681
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.00%	Indels:	0
DB:	17	Gaps:	0

US-10-799-747-116 (1-20) x US-10-322-281-411 (1-66681)

```
Qy          4 HisSerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||||
Db      38196 CACTCTGTTCTTTCATTTCTTTTG 38219
```

RESULT 10
US-10-322-281-858/c
; Sequence 858, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 858
; LENGTH: 133787
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(133787)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-858

Alignment Scores:

Pred. No.:	2.61e+03	Length:	133787
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.00%	Indels:	0
DB:	17	Gaps:	0

US-10-799-747-116 (1-20) x US-10-322-281-858 (1-133787)

```
Qy          4 HisSerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||||
Db      35487 CACTCAGTTCTATCATTTCTTCTT 35464
```

RESULT 11
US-10-322-281-612
; Sequence 612, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 612
; LENGTH: 196063
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(196063)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-612

Alignment Scores:

Pred. No.:	3.7e+03	Length:	196063
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.00%	Indels:	0
DB:	17	Gaps:	0

US-10-799-747-116 (1-20) x US-10-322-281-612 (1-196063)

```
Qy          3 AlaHisSerValLeuSerPheLeu 10
              ||||||||||||||||||||||||
Db      50669 GCTCATTCTGTACTTTCCTTCCTC 50692

RESULT 12
US-10-027-632-59064/c
; Sequence 59064, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 59064
;   LENGTH: 1601042
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1601042)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59064

Alignment Scores:
Pred. No.:                2.55e+04        Length:          1601042
Score:                    8.00            Matches:         8
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              40.00%          Indels:          0
DB:                       13              Gaps:            0

US-10-799-747-116 (1-20) x US-10-027-632-59064 (1-1601042)

Qy          3 AlaHisSerValLeuSerPheLeu 10
              ||||||||||||||||||||||||
Db     638163 GCCCACTCGGTCCTTAGCTTTCTG 638140

RESULT 13
US-10-027-632-59064/c
; Sequence 59064, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 59064
;   LENGTH: 1601042
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1601042)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59064

Alignment Scores:
Pred. No.:                2.55e+04        Length:          1601042
Score:                    8.00            Matches:         8
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              40.00%          Indels:          0
DB:                       16              Gaps:            0

US-10-799-747-116 (1-20) x US-10-027-632-59064 (1-1601042)

Qy          3 AlaHisSerValLeuSerPheLeu 10
              ||||||||||||||||||||||||
Db     638163 GCCCACTCGGTCCTTAGCTTTCTG 638140

RESULT 14
US-10-032-585-3022
; Sequence 3022, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3022
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-10-032-585-3022

Alignment Scores:
Pred. No.:                26.1            Length:          65
Score:                    7.00            Matches:         7
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              35.00%          Indels:          0
DB:                       15              Gaps:            0

US-10-799-747-116 (1-20) x US-10-032-585-3022 (1-65)

Qy          5 SerValLeuSerPheLeuLeu 11
              |||||||||||||||||||
Db         13 TCTGTATTATCATTCTTATTG 33

RESULT 15
US-10-424-599-127928
; Sequence 127928, Application US/10424599
; Publication No. US20040031072A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 127928
;   LENGTH: 163
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_86523C.1
US-10-424-599-127928

Alignment Scores:
Pred. No.:                60.8          Length:          163
Score:                    7.00          Matches:         7
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              35.00%        Indels:          0
DB:                       13            Gaps:            0

US-10-799-747-116 (1-20) x US-10-424-599-127928 (1-163)

Qy          4 HisSerValLeuSerPheLeu 10
              ||||||||||||||||||||
Db         40 CACTCTGTGTTGTCCTTTCTG 60


Search completed: July 21, 2004, 07:54:36
Job time : 3580 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 03:49:02 ; Search time 2467 Seconds
(without alignments)
242.093 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 20
Sequence: 1 MAAHSVLSFLLWTPYALKSX 20

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55023894

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10799747/runat_19072004_161443_21994/app_query.fasta_1.199
-DB=EST -QFMT=fastap -SUFFIX=olig.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747_@CGN_1_1_5180_@runat_19072004_161443_21994 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	19	95.0	240	10	BF910533	BF910533 CM4-UT004
2	19	95.0	575	12	BM722991	BM722991 UI-E-EO1-
3	19	95.0	765	12	BI914473	BI914473 603182264
4	19	95.0	785	12	BG484396	BG484396 602505037
5	19	95.0	1201	13	BX441923	BX441923 BX441923
6	10	50.0	621	9	AL035941	AL035941 DKFZp564B
7	9	45.0	496	28	AQ293004	AQ293004 HS_2225_A
8	8	40.0	239	9	AV851019	AV851019 AV851019
9	8	40.0	272	12	BG371993	BG371993 UI-R-CV0-
c 10	8	40.0	290	29	CE488576	CE488576 tigr-gss-
c 11	8	40.0	323	28	AQ350287	AQ350287 RPCI11-13
c 12	8	40.0	328	29	CC808823	CC808823 ZMMBBc047
13	8	40.0	336	14	CD202965	CD202965 MS1-0139P
14	8	40.0	356	14	T59368	T59368 yb57h01.s1
15	8	40.0	385	29	CE286752	CE286752 tigr-gss-
16	8	40.0	396	28	BH363737	BH363737 CH230-127
c 17	8	40.0	400	28	AQ475321	AQ475321 CITBI-E1-
c 18	8	40.0	419	13	BW204790	BW204790 BW204790
19	8	40.0	430	14	CD201054	CD201054 MS1-0124U
20	8	40.0	430	14	CD201131	CD201131 MS1-0124U
21	8	40.0	441	9	AV888988	AV888988 AV888988
22	8	40.0	444	10	BF412147	BF412147 UI-R-BT1-
23	8	40.0	448	14	CD113088	CD113088 ME1-0028T
24	8	40.0	448	14	CD201074	CD201074 MS1-0124U
25	8	40.0	448	14	CD201080	CD201080 MS1-0124U
26	8	40.0	448	14	CD201099	CD201099 MS1-0124U
27	8	40.0	448	14	CD201120	CD201120 MS1-0124U
c 28	8	40.0	457	9	AV981198	AV981198 AV981198
c 29	8	40.0	458	28	BZ615844	BZ615844 ig55d08.b
30	8	40.0	468	28	AQ945119	AQ945119 Sheared D
31	8	40.0	470	28	AQ702871	AQ702871 HS_5452_B
c 32	8	40.0	481	13	BW196597	BW196597 BW196597
33	8	40.0	490	9	AV890393	AV890393 AV890393
34	8	40.0	493	9	AV862905	AV862905 AV862905
c 35	8	40.0	501	13	BW202753	BW202753 BW202753
c 36	8	40.0	504	14	CD112744	CD112744 ME1-0024T
c 37	8	40.0	505	14	CA820687	CA820687 EST0012 N
c 38	8	40.0	510	9	AV950469	AV950469 AV950469
39	8	40.0	510	29	TA90C02P	AL465308 T. brucei
c 40	8	40.0	515	29	CC476294	CC476294 CH240_302
c 41	8	40.0	516	12	BP001077	BP001077 BP001077
42	8	40.0	521	13	BW136624	BW136624 BW136624
c 43	8	40.0	524	13	BW203262	BW203262 BW203262
44	8	40.0	525	9	AV680124	AV680124 AV680124
c 45	8	40.0	526	9	AV973566	AV973566 AV973566

ALIGNMENTS

RESULT 1
BF910533/c
LOCUS BF910533 240 bp mRNA linear EST 18-JAN-2001
DEFINITION CM4-UT0042-011100-395-f08 UT0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF910533
VERSION BF910533.1 GI:12301991
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 240)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

```
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM4&t2=CM4-UT0042-
            011100-395-f08&t3=2000-11-01&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 81
            High quality sequence stop: 240.
FEATURES             Location/Qualifiers
     source          1. .240
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="UT0042"
                     /note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

Alignment Scores:
Pred. No.:                6.33e-09        Length:          240
Score:                    19.00           Matches:         19
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              95.00%          Indels:          0
DB:                       10              Gaps:            0

US-10-799-747-116 (1-20) x BF910533 (1-240)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 135

RESULT 2
BM722991
LOCUS       BM722991                 575 bp    mRNA    linear   EST 01-MAR-2002
DEFINITION  UI-E-EO1-aid-a-14-0-UI.r1 UI-E-EO1 Homo sapiens cDNA clone
            UI-E-EO1-aid-a-14-0-UI 5', mRNA sequence.
ACCESSION   BM722991
VERSION     BM722991.1  GI:19043963
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 148-169, >AT_rich#Low_complexity (matched compliment)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EO1-aid-a-14-0-UI"
                     /tissue_type="fetal eye"
                     /dev_stage="fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-EO1"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EO1 is a normalized cDNA library containing the
                     following tissue(s): fetal eye. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     CGCGTATACC. This library was created for the program, Gene
                     Discovery in the Visual System, supported by National Eye
                     Institute (NEI)."
ORIGIN

Alignment Scores:
Pred. No.:                1.55e-08        Length:          575
Score:                    19.00           Matches:         19
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              95.00%          Indels:          0
DB:                       12              Gaps:            0

US-10-799-747-116 (1-20) x BM722991 (1-575)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 510

RESULT 3
BI914473
LOCUS       BI914473                 765 bp    mRNA    linear   EST 16-OCT-2001
DEFINITION  603182264F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246463 5',
            mRNA sequence.
ACCESSION   BI914473
VERSION     BI914473.1  GI:16178652
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
```

```
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11621  row: p  column: 16
            High quality sequence stop: 762.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5246463"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_121"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
                     fetal brains, female age 20 weeks, female age 24 weeks,
                     and male age 26 weeks.  Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     0.7-3.5 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 017. Note:
                     this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                2.07e-08        Length:         765
Score:                    19.00           Matches:        19
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              95.00%          Indels:         0
DB:                       12              Gaps:           0

US-10-799-747-116 (1-20) x BI914473 (1-765)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 612

RESULT 4
BG484396
LOCUS       BG484396                 785 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602505037F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618473 5',
            mRNA sequence.
ACCESSION   BG484396
VERSION     BG484396.1  GI:13416675
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 785)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1376  row: j  column: 10
            High quality sequence start: 18
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1. .785
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4618473"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_77"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                2.13e-08        Length:         785
Score:                    19.00           Matches:        19
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              95.00%          Indels:         0
DB:                       12              Gaps:           0

US-10-799-747-116 (1-20) x BG484396 (1-785)

Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 378

RESULT 5
BX441923
LOCUS       BX441923                1201 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX441923 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF023YO02 5-PRIME, mRNA sequence.
ACCESSION   BX441923
VERSION     BX441923.1  GI:30771989
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2532.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF023BH01QP1&cluster=2532.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF023BH01QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF023YO02"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
```

/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:

Pred. No.:	3.28e-08	Length:	1201
Score:	19.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	95.00%	Indels:	0
DB:	13	Gaps:	0

US-10-799-747-116 (1-20) x BX441923 (1-1201)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrpThrProTyrAlaLeuLysSer 19
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 ATGGCAGCCCATTCAGTCTTGAGTTTTCTTCTCTGGACACCTTATGCTCTGAAATCA 202
```

RESULT 6
AL035941
LOCUS AL035941 621 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp564B1622_r1 564 (synonym: hfbr2) Homo sapiens cDNA clone DKFZp564B1622_5', mRNA sequence.
ACCESSION AL035941
VERSION AL035941.1 GI:5405575
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 621)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Wambutt, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
s1 sequence also available.
This clone (DKFZp564B1622) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES Location/Qualifiers
source 1. .621
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp564B1622"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="Xl-2blue"
/clone_lib="564 (synonym: hfbr2)"
/note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"
ORIGIN

Alignment Scores:

Pred. No.:	6.57	Length:	621
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	50.00%	Indels:	0
DB:	9	Gaps:	0

US-10-799-747-116 (1-20) x AL035941 (1-621)

```
Qy          1 MetAlaAlaHisSerValLeuSerPheLeu 10
              ||||||||||||||||||||||||||||||
Db        591 ATGGCAGCCCATTCAGTCTTGAGTTTTCTT 620
```

RESULT 7
AQ293004
LOCUS AQ293004 496 bp DNA linear GSS 15-DEC-1998
DEFINITION HS_2225_A1_E02_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2225 Col=3 Row=I, genomic survey sequence.
ACCESSION AQ293004
VERSION AQ293004.1 GI:4010727
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 496)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2225 row: I column: 3
Class: BAC ends
High quality sequence stop: 496.
FEATURES Location/Qualifiers
source 1. .496
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=2225 Col=3 Row=I"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
ORIGIN

Alignment Scores:

Pred. No.:	47.1	Length:	496
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.00%	Indels:	0
DB:	28	Gaps:	0

US-10-799-747-116 (1-20) x AQ293004 (1-496)

```
Qy          5 SerValLeuSerPheLeuLeuTrpThr 13
              |||||||||||||||||||||||||
Db        119 TCTGTATTATCCTTCCTGCTATGGACT 145
```

RESULT 8
AV851019
LOCUS AV851019 239 bp mRNA linear EST 08-NOV-2001
DEFINITION AV851019 Nori Satoh unpublished cDNA library, larva Ciona intestinalis cDNA clone rcilv10n02 3', mRNA sequence.
ACCESSION AV851019
VERSION AV851019.1 GI:16834788
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

```
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 239)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .239
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="rcilv10n02"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /clone_lib="Nori Satoh unpublished cDNA library, larva"
ORIGIN

Alignment Scores:
Pred. No.:                 201          Length:         239
Score:                     8.00         Matches:        8
Percent Similarity:        100.00%      Conservative:   0
Best Local Similarity:     100.00%      Mismatches:     0
Query Match:               40.00%       Indels:         0
DB:                        9            Gaps:           0

US-10-799-747-116 (1-20) x AV851019 (1-239)

Qy          4 HisSerValLeuSerPheLeuLeu 11
              ||||||||||||||||||||||||
Db        173 CATTCAGTTTTGTCTTTTTTGCTC 196

RESULT 9
BG371993
LOCUS       BG371993                272 bp    mRNA    linear   EST 09-MAR-2001
DEFINITION  UI-R-CV0-brl-a-01-0-UI.s1 UI-R-CV0 Rattus norvegicus cDNA clone
            UI-R-CV0-brl-a-01-0-UI 3', mRNA sequence.
ACCESSION   BG371993
VERSION     BG371993.1  GI:13268530
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 272)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized rat eye library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 1-25,
            >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .272
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CV0-brl-a-01-0-UI"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-CV0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CV0
                     library is a non-normalized library constructed from rat
                     eye tissue. For a detailed description of the library from
                     which this clone was derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_TISSUE=rat eye
                     TAG_LIB=UI-R-CV0
                     TAG_SEQ=CAGCC"
ORIGIN

Alignment Scores:
Pred. No.:                 230          Length:         272
Score:                     8.00         Matches:        8
Percent Similarity:        100.00%      Conservative:   0
Best Local Similarity:     100.00%      Mismatches:     0
Query Match:               40.00%       Indels:         0
DB:                        12           Gaps:           0

US-10-799-747-116 (1-20) x BG371993 (1-272)

Qy          2 AlaAlaHisSerValLeuSerPhe 9
              |||||||||||||||||||||||
Db        107 GCAGCACATTCTGTGCTTTCCTTC 130

RESULT 10
CE488576/c
LOCUS       CE488576                290 bp    DNA     linear   GSS 28-SEP-2003
DEFINITION  tigr-gss-dog-17000365567544 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE488576
VERSION     CE488576.1  GI:36805357
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 290)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .290
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
```

```
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

Alignment Scores:
Pred. No.:                245             Length:         290
Score:                    8.00            Matches:        8
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              40.00%          Indels:         0
DB:                       29              Gaps:           0

US-10-799-747-116 (1-20) x CE488576 (1-290)

Qy          4 HisSerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db         43 CACTCTGTCCTATCGTTTCTACTG 20

RESULT 11
AQ350287/c
LOCUS       AQ350287                323 bp    DNA     linear   GSS 07-MAY-1999
DEFINITION  RPCI11-133N7.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-133N7,
            genomic survey sequence.
ACCESSION   AQ350287
VERSION     AQ350287.1  GI:4177622
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7551006"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-133N7"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

Alignment Scores:
Pred. No.:                274             Length:         323
Score:                    8.00            Matches:        8
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              40.00%          Indels:         0
DB:                       28              Gaps:           0

US-10-799-747-116 (1-20) x AQ350287 (1-323)

Qy          4 HisSerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||
Db        137 CATTCCGTCCTCTCCTTCCTTCTT 114

RESULT 12
CC808823/c
LOCUS       CC808823                328 bp    DNA     linear   GSS 14-JUL-2003
DEFINITION  ZMMBBc0476P22r ZMMBBc Zea mays subsp. mays genomic clone
            ZMMBBc0476P22 3', genomic survey sequence.
ACCESSION   CC808823
VERSION     CC808823.1  GI:32641958
KEYWORDS    GSS.
SOURCE      Zea mays subsp. mays (maize)
  ORGANISM  Zea mays subsp. mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Rouzard,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003b)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: SP6
            Class: BAC ends
            High quality sequence start: 88.
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Zea mays subsp. mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /sub_species="mays"
                     /db_xref="taxon:4578"
                     /clone="ZMMBBc0476P22"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBc"
                     /note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"
ORIGIN

Alignment Scores:
Pred. No.:                278             Length:         328
Score:                    8.00            Matches:        8
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              40.00%          Indels:         0
DB:                       29              Gaps:           0

US-10-799-747-116 (1-20) x CC808823 (1-328)

Qy          3 AlaHisSerValLeuSerPheLeu 10
              |||||||||||||||||||||
Db         35 GCTCACTCTGTCCTTAGTTTTTTG 12

RESULT 13
CD202965
LOCUS       CD202965                336 bp    mRNA    linear   EST 14-SEP-2003
DEFINITION  MS1-0139P-V386-G01-U.B MS1-0139 Schistosoma mansoni cDNA clone
            MS1-0139P-V386-G01.B, mRNA sequence.
ACCESSION   CD202965
VERSION     CD202965.1  GI:34732656
KEYWORDS    EST.
SOURCE      Schistosoma mansoni
```

```
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
            Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
            Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
            Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
            Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
            Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
            Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
            Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
            Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
  TITLE     Transcriptome analysis of the acoelomate human parasite Schistosoma
            mansoni
  JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
  MEDLINE   22879926
COMMENT     Contact: Dr. Sergio Verjovski-Almeida
            Departamento de Bioquimica
            Instituto de Quimica - Universidade de Sao Paulo
            Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
            Brasil
            Tel: +55-11-3091-2173
            Fax: +55-11-3091-2186
            Email: verjo@iq.usp.br
            This sequence was derived from the FAPESP Schistosoma mansoni EST
            Genome Project. All sequences in the project were assembled and
            annotated. This entry and all the assembled sequences can be seen
            in the following URL http://bioinfo.iq.usp.br/schisto/
            Plate: MS1-0139P-V386  row: 1  column: G.
FEATURES                Location/Qualifiers
     source             1. .336
                        /organism="Schistosoma mansoni"
                        /mol_type="mRNA"
                        /db_xref="taxon:6183"
                        /clone="MS1-0139P-V386-G01.B"
                        /sex="mixed pool"
                        /dev_stage="schistosomulum"
                        /lab_host="in vitro culture"
                        /clone_lib="MS1-0139"
                        /note="Vector: pGEM T-easy"
ORIGIN

Alignment Scores:
Pred. No.:                285             Length:         336
Score:                    8.00            Matches:        8
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              40.00%          Indels:         0
DB:                       14              Gaps:           0

US-10-799-747-116 (1-20) x CD202965 (1-336)

Qy          4 HisSerValLeuSerPheLeuLeu 11
              |||||||||||||||||||||||||
Db         99 CATTCGGTTTTAAGTTTTCTTCTG 122

RESULT 14
T59368
LOCUS       T59368                  356 bp    mRNA    linear   EST 09-FEB-1995
DEFINITION  yb57h01.s1 Stratagene ovary (#937217) Homo sapiens cDNA clone
            IMAGE:75313 3' similar to gb:X77738_rna1 BAND 3 ANION TRANSPORT
            PROTEIN (HUMAN);contains Alu repetitive element;, mRNA sequence.
ACCESSION   T59368
VERSION     T59368.1  GI:661205
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
  PUBMED    8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 339
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -21m13
            High quality sequence stop: 339.
FEATURES                Location/Qualifiers
     source             1. .356
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="GDB:505042"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:75313"
                        /sex="female"
                        /dev_stage="49 year old"
                        /lab_host="SOLR cells (kanamycin resistant)"
                        /clone_lib="Stratagene ovary (#937217)"
                        /note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
                        Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                        Total ovary tissue, normal, caucasian. Average insert
                        size: 0.8 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
                        GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                        CTCGAGTTTTTTTTTTTTTTTT 3'"
ORIGIN

Alignment Scores:
Pred. No.:                302             Length:         356
Score:                    8.00            Matches:        8
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              40.00%          Indels:         0
DB:                       14              Gaps:           0

US-10-799-747-116 (1-20) x T59368 (1-356)

Qy          3 AlaHisSerValLeuSerPheLeu 10
              ||||||||||||||||||||||||
Db        108 GCTCATTCTGTACTTTCCTTCCTC 131

RESULT 15
CE286752
LOCUS       CE286752                385 bp    DNA     linear   GSS 26-SEP-2003
DEFINITION  tigr-gss-dog-17000333810288 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE286752
VERSION     CE286752.1  GI:36056365
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
```

```
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

Alignment Scores:
Pred. No.:              327           Length:         385
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            40.00%        Indels:         0
DB:                     29            Gaps:           0

US-10-799-747-116 (1-20) x CE286752 (1-385)

Qy          4 HisSerValLeuSerPheLeuLeu 11
              ||||||||||||||||||||||
Db        198 CACAGCGTCTTATCATTTCTCTTG 221


Search completed: July 21, 2004, 05:22:32
Job time : 2475 secs
```